

US-09-020-116-4

Query Match 0.9%; Score 7; DB 3; Length 27;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 GKDSLSD 563
| | | | |
Db 8 GKDSLSD 14

RESULT 25

US-09-608-902-4
; Sequence 4, Application US/09608902
; Patent No. 6353097
; GENERAL INFORMATION:
; APPLICANT: Vonakis, Becky
; APPLICANT: Metzger, Henry
; APPLICANT: Chen, Huaxian
; TITLE OF INVENTION: SIGNAL TRANSDUCTION INHIBITORS OF
; TITLE OF INVENTION: ALLERGIC REACTIONS
; FILE REFERENCE: 14014.028502
; CURRENT APPLICATION NUMBER: US/09/608,902
; CURRENT FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 09/020,116
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: /No. 6353097e = synthetic cons
US-09-608-902-4

Query Match 0.9%; Score 7; DB 4; Length 27;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 GKDSLSD 563
| | | | |
Db 8 GKDSLSD 14

Search completed: October 1, 2004, 07:35:19
Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 07:33:31 ; Search time 90 Seconds
(without alignments)
2846.135 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 796

Sequence: 1 SYELGLYQARTVKNENRVS.....KLLALLKGNPSSVSKKIN 796

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1351062 seqs, 321799191 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1351062

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	795	99.9	796	9	US-09-765-272-56
2	695	87.3	819	12	US-10-412-862-8
3	695	87.3	819	15	US-10-412-850-8
4	695	87.3	819	15	US-10-387-783-8
5	80	10.1	334	15	US-10-324-143-20
6	80	10.1	690	15	US-10-324-143-38
7	80	10.1	821	15	US-10-324-143-19
8	80	10.1	840	10	US-09-884-465A-7
9	80	10.1	840	15	US-10-324-143-8
10	60	7.5	819	12	US-10-412-862-10
11	60	7.5	819	15	US-10-324-143-32
12	60	7.5	819	15	US-10-412-850-10
13	60	7.5	819	15	US-10-387-783-10
14	60	7.5	826	10	US-09-769-787-194
15	60	7.5	838	10	US-09-884-465A-8

16	60	7.5	838	12	US-10-412-862-4	Sequence 4, Appli
17	60	7.5	838	15	US-10-324-143-9	Sequence 9, Appli
18	60	7.5	838	15	US-10-412-850-4	Sequence 4, Appli
19	60	7.5	838	15	US-10-387-783-4	Sequence 4, Appli
20	56	7.0	613	15	US-10-324-143-22	Sequence 22, Appli
21	51	6.4	763	9	US-09-765-272-66	Sequence 66, Appli
22	32	4.0	42	10	US-09-884-465A-12	Sequence 12, Appli
23	32	4.0	205	15	US-10-324-143-18	Sequence 18, Appli
24	32	4.0	447	9	US-09-765-272-182	Sequence 182, App
25	32	4.0	484	10	US-09-769-787-38	Sequence 38, Appli
26	32	4.0	484	10	US-09-769-744A-24	Sequence 24, Appli
27	32	4.0	484	12	US-10-412-862-6	Sequence 6, Appli
28	32	4.0	484	15	US-10-412-850-6	Sequence 6, Appli
29	32	4.0	484	15	US-10-387-783-6	Sequence 6, Appli
30	32	4.0	489	15	US-10-324-143-15	Sequence 15, Appli
31	32	4.0	509	15	US-10-324-143-16	Sequence 16, Appli
32	32	4.0	780	15	US-10-324-143-37	Sequence 37, Appli
33	32	4.0	840	10	US-09-884-465A-10	Sequence 10, Appli
34	32	4.0	1019	15	US-10-324-143-14	Sequence 14, Appli
35	32	4.0	1039	10	US-09-884-465A-6	Sequence 6, Appli
36	32	4.0	1039	15	US-10-324-143-7	Sequence 7, Appli
37	27	3.4	294	15	US-10-324-143-39	Sequence 39, Appli
38	27	3.4	473	15	US-10-324-143-36	Sequence 36, Appli
39	27	3.4	568	15	US-10-324-143-33	Sequence 33, Appli
40	27	3.4	612	15	US-10-324-143-44	Sequence 44, Appli
41	27	3.4	1139	10	US-09-884-465A-380	Sequence 380, App
42	27	3.4	1139	15	US-10-324-143-51	Sequence 51, Appli
43	27	3.4	1378	10	US-09-884-465A-378	Sequence 378, App
44	27	3.4	1378	15	US-10-324-143-49	Sequence 49, Appli
45	24	3.0	226	15	US-10-324-143-30	Sequence 30, Appli
46	24	3.0	428	15	US-10-324-143-27	Sequence 27, Appli
47	24	3.0	555	15	US-10-324-143-26	Sequence 26, Appli
48	24	3.0	999	10	US-09-884-465A-376	Sequence 376, App
49	24	3.0	999	10	US-09-884-465A-377	Sequence 377, App
50	24	3.0	999	15	US-10-324-143-47	Sequence 47, Appli
51	24	3.0	999	15	US-10-324-143-48	Sequence 48, Appli
52	24	3.0	1126	10	US-09-884-465A-383	Sequence 383, App
53	24	3.0	1126	15	US-10-324-143-54	Sequence 54, Appli
54	24	3.0	1238	10	US-09-884-465A-381	Sequence 381, App
55	24	3.0	1238	15	US-10-324-143-52	Sequence 52, Appli
56	24	3.0	1365	10	US-09-884-465A-382	Sequence 382, App
57	24	3.0	1365	15	US-10-324-143-53	Sequence 53, Appli
58	23	2.9	94	10	US-09-884-465A-11	Sequence 11, Appli
59	17	2.1	381	10	US-09-769-736-24	Sequence 24, Appli
60	17	2.1	793	10	US-09-252-088-15	Sequence 15, Appli
61	17	2.1	793	15	US-10-340-792-18	Sequence 18, Appli
62	17	2.1	822	10	US-09-769-736-18	Sequence 18, Appli
63	15	1.9	1058	15	US-10-324-143-46	Sequence 46, Appli
64	14	1.8	487	15	US-10-324-143-21	Sequence 21, Appli
65	14	1.8	487	15	US-10-324-143-35	Sequence 35, Appli
66	13	1.6	132	15	US-10-324-143-29	Sequence 29, Appli
67	12	1.5	203	15	US-10-324-143-31	Sequence 31, Appli
68	12	1.5	214	15	US-10-324-143-42	Sequence 42, Appli
69	12	1.5	342	15	US-10-324-143-40	Sequence 40, Appli
70	12	1.5	715	10	US-09-252-088-16	Sequence 16, Appli
71	12	1.5	715	15	US-10-340-792-16	Sequence 16, Appli
72	12	1.5	807	15	US-10-324-143-45	Sequence 45, Appli
73	12	1.5	913	10	US-09-884-465A-384	Sequence 384, App
74	12	1.5	913	15	US-10-324-143-55	Sequence 55, Appli
75	12	1.5	1152	10	US-09-884-465A-379	Sequence 379, App
76	12	1.5	1152	15	US-10-324-143-50	Sequence 50, Appli
77	11	1.4	485	10	US-09-769-736-72	Sequence 72, Appli
78	9	1.1	121	15	US-10-324-143-28	Sequence 28, Appli
79	9	1.1	124	12	US-10-424-599-256231	Sequence 256231,
80	9	1.1	272	15	US-09-884-465A-300	Sequence 300, App
81	9	1.1	272	15	US-10-324-143-85	Sequence 85, Appli
82	9	1.1	782	12	US-10-282-122A-44621	Sequence 44621, A
83	9	1.1	791	12	US-10-282-122A-67300	Sequence 67300, A
84	8	1.0	86	10	US-09-884-465A-22	Sequence 22, Appli
85	8	1.0	215	16	US-10-389-566-597	Sequence 597, App
86	8	1.0	257	15	US-10-369-493-9660	Sequence 9660, Ap
87	8	1.0	259	10	US-09-884-465A-374	Sequence 374, App
88	8	1.0	270	10	US-09-884-465A-302	Sequence 302, App

Sequence 87, Appl
Sequence 258, App

8 1.0 270 15 US-10-324-143-87
90 1.0 272 10 US-09-884-465A-258

ALIGNMENTS

RESULT 1
US-09-765-272-56
; Sequence 56, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 796 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 56:

US-09-765-272-56
Query Match 99.9%; Score 795; DB 9; Length 796;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKNRRVSYIDGKQATKTENLTDPVSKREGINAEQIVIKITDQGYVT 60

Db 1 SYELGLYQARTVKNRRVSYIDGKQATKTENLTDPVSKREGINAEQIVIKITDQGYVT 60

QY 61 SHGDHYHYNGKVPDVAIISEBLLMKDPNYKLDKEDIQVNEVGKGVYKVDGKYVYLKDA 120

Db 61 SHGDHYHYNGKVPDVAIISEBLLMKDPNYKLDKEDIQVNEVGKGVYKVDGKYVYLKDA 120

QY 121 AHADNVRTKEENRQKQESQREHGTGPRNDGVALARSQGRYTTDDGYIFNASDIIEDT 180

Db 121 AHADNVRTKEENRQKQESQREHGTGPRNDGVALARSQGRYTTDDGYIFNASDIIEDT 180

QY 181 GDAYIVPHGDHYHYIPKNELSAELAAAEAFISGRGNLSNRTYRQNSDNTSRTNWVPS 240

Db 181 GDAYIVPHGDHYHYIPKNELSAELAAAEAFISGRGNLSNRTYRQNSDNTSRTNWVPS 240

QY 241 VSNPGTTNTNTNNTNSQASQSDNDISLLKQLKPLSORHVESDGLVFPDPAITSR 300

Db 241 VSNPGTTNTNTNNTNSQASQSDNDISLLKQLKPLSORHVESDGLVFPDPAITSR 300

Db 241 VSNPGTTNTNTNNTNSQASQSDNDISLLKQLKPLSORHVESDGLVFPDPAITSR 300

QY 301 ARGVAVPHGDHYHYIPYSQMSLEERIIARIIPLRVSRNHWVDPDSRPEQSPQPTPEPSPG 360

Db 301 ARGVAVPHGDHYHYIPYSQMSLEERIIARIIPLRVSRNHWVDPDSRPEQSPQPTPEPSPG 360

QY 361 PQAPNLKIDNSSLSVQLVRKVGSGYVFEKGISSRYVFAKDLPSSETVKNLESKLSKQES 420

Db 361 PQAPNLKIDNSSLSVQLVRKVGSGYVFEKGISSRYVFAKDLPSSETVKNLESKLSKQES 420

QY 421 VSHTLTAKKENVAPRDQBFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480

Db 421 VSHTLTAKKENVAPRDQBFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480

QY 481 KEKLVDDLLAFAPITTHPERLGKPNQJTEYDEVRIAQLADKYTTSDGYIFDEHDIISD 540

Db 481 KEKLVDDLLAFAPITTHPERLGKPNQJTEYDEVRIAQLADKYTTSDGYIFDEHDIISD 540

QY 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAAYTEKGIILPPSPDADVKANPTGDSAAAIY 600

Db 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAAYTEKGIILPPSPDADVKANPTGDSAAAIY 600

QY 601 NRKGEKRIPLVRLPYMVEHTVEKNGNLIIPHKKDHYHNIKFAWEDDHTYKAPNGYTLED 660

Db 601 NRKGEKRIPLVRLPYMVEHTVEKNGNLIIPHKKDHYHNIKFAWEDDHTYKAPNGYTLED 660

QY 661 LPATIKYVVEHPDERPHSNDGWNASEHVLGKHSDPNKNFKADEPVEETPAEPEVP 720

Db 661 LPATIKYVVEHPDERPHSNDGWNASEHVLGKHSDPNKNFKADEPVEETPAEPEVP 720

QY 721 QVETEKVEAQLKEAEVLLAKVTDSSLKANATETTLAQLRNNLTQIMDNNSIMAEAEKLLA 780

Db 721 QVETEKVEAQLKEAEVLLAKVTDSSLKANATETTLAQLRNNLTQIMDNNSIMAEAEKLLA 780

QY 781 LLKGSNPSVSKEKIN 796

Db 781 LLKGSNPSVSKEKIN 796

RESULT 2
US-10-412-862-8
; Sequence 8, Application US/10412862
; Publication No. US20040052781A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-685
; CURRENT APPLICATION NUMBER: US/10/412,862
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 8
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-412-862-8

Query Match 87.3%; Score 695; DB 12; Length 819;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKNRRVSYIDGKQATKTENLTDPVSKREGINAEQIVIKITDQGYVT 60

Db 21 SYELGLYQARTVKNRRVSYIDGKQATKTENLTDPVSKREGINAEQIVIKITDQGYVT 80

QY 61 SHGDHYHYNGKVPDVAIISEBLLMKDPNYKLDKEDIQVNEVGKGVYKVDGKYVYLKDA 120


```

Db      81  SHGDHYYNGKVPYDAIISEELLMKDPNKLKDEDI VNEVKGYIVKVDGKYVYVLKDA 140
QY      121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSQRYTTDDGYIFNADSIIEDT 180
Db      141 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSQRYTTDDGYIFNADSIIEDT 200
QY      181 GDAYIVPHGDHYHYIIPKNELASASELAFLSGRGNLSNRYRONSNTSRTNWVPS 240
Db      201 GDAYIVPHGDHYHYIIPKNELASASELAFLSGRGNLSNRYRONSNTSRTNWVPS 260
QY      241 VSNPGTTNTNTSNNSTNSOASQSDNDISLLKQYKLPISQRHVESDGLVDFDPAQITTSRT 300
Db      261 VSNPGTTNTNTSNNSTNSOASQSDNDISLLKQYKLPISQRHVESDGLVDFDPAQITTSRT 320
QY      301 ARGVAVPHGDHYHYIIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQTPPEPSG 360
Db      321 ARGVAVPHGDHYHYIIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQTPPEPSG 380
QY      361 PQAPNLKIDSNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSSETVKNLESKLSKQES 420
Db      381 PQAPNLKIDSNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSSETVKNLESKLSKQES 440
QY      421 VSHLTAKKENVAPRQDEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
Db      441 VSHLTAKKENVAPRQDEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 500
QY      481 KEKLVDDLAFAPITHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDEHDIISD 540
Db      501 KEKLVDDLAFAPITHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDEHDIISD 560
QY      541 EGDAYVTPHMGHSHWIGKDSLSKDKVAAQAYTKEKGILPPSPDADVKANPTGDSAAIY 600
Db      561 EGDAYVTPHMGHSHWIGKDSLSKDKVAAQAYTKEKGILPPSPDADVKANPTGDSAAIY 620
QY      601 NRVKGEKRIPLVRLPYMVEHTVEKNGNLIIPHKHVHNIKFAWFDHDTYKAPNGYTTLED 660
Db      621 NRVKGEKRIPLVRLPYMVEHTVEKNGNLIIPHKHVHNIKFAWFDHDTYKAPNGYTTLED 680
QY      661 LFATIKYVVEHPDHPHSDNGWGNASEHVLGKDKHSDPNKNFKADPEPVEETPAEPEVP 720
Db      681 LFATIKYVVEHPDHPHSDNGWGNASEHVLGKDKHSDPNKNFKADPEPVEETPAEPEVP 740
QY      721 QVETEKVEAQLKEAEVLLAKVTSSSKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 780
Db      741 QVETEKVEAQLKEAEVLLAKVTSSSKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 800
QY      781 LLKGSNPSSVSKEKIN 796
Db      801 LLKGSNPSSVSKEKIN 816

```

RESULT 3

```

US-10-412-850-8
; Sequence 8, Application US/10412850
; Publication No. US20040001836A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-686
; CURRENT APPLICATION NUMBER: US/10/412,850
; PRIOR FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 8
; LENGTH: 819

```

RESULT 4

```

US-10-387-783-8
; Sequence 8, Application US/10387783
; Publication No. US20040005331A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus

```

```

; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-412-850-8

```

```

Query Match      87.3%; Score 695; DB 15; Length 819;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1  SYELGLYQARTVKNRVSVIDKQATQKTENITPDEVSKREGINAEQIVIKITDQGYVT 60
Db      21  SYELGLYQARTVKNRVSVIDKQATQKTENITPDEVSKREGINAEQIVIKITDQGYVT 80
QY      61  SHGDHYYNGKVPYDAIISEELLMKDPNKLKDEDI VNEVKGYIVKVDGKYVYVLKDA 120
Db      81  SHGDHYYNGKVPYDAIISEELLMKDPNKLKDEDI VNEVKGYIVKVDGKYVYVLKDA 140
QY      121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSQRYTTDDGYIFNADSIIEDT 180
Db      141 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSQRYTTDDGYIFNADSIIEDT 200
QY      181 GDAYIVPHGDHYHYIIPKNELASASELAFLSGRGNLSNRYRONSNTSRTNWVPS 240
Db      201 GDAYIVPHGDHYHYIIPKNELASASELAFLSGRGNLSNRYRONSNTSRTNWVPS 260
QY      241 VSNPGTTNTNTSNNSTNSOASQSDNDISLLKQYKLPISQRHVESDGLVDFDPAQITTSRT 300
Db      261 VSNPGTTNTNTSNNSTNSOASQSDNDISLLKQYKLPISQRHVESDGLVDFDPAQITTSRT 320
QY      301 ARGVAVPHGDHYHYIIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQTPPEPSG 360
Db      321 ARGVAVPHGDHYHYIIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQTPPEPSG 380
QY      361 PQAPNLKIDSNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSSETVKNLESKLSKQES 420
Db      381 PQAPNLKIDSNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSSETVKNLESKLSKQES 440
QY      421 VSHLTAKKENVAPRQDEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
Db      441 VSHLTAKKENVAPRQDEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 500
QY      481 KEKLVDDLAFAPITHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDEHDIISD 540
Db      501 KEKLVDDLAFAPITHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDEHDIISD 560
QY      541 EGDAYVTPHMGHSHWIGKDSLSKDKVAAQAYTKEKGILPPSPDADVKANPTGDSAAIY 600
Db      561 EGDAYVTPHMGHSHWIGKDSLSKDKVAAQAYTKEKGILPPSPDADVKANPTGDSAAIY 620
QY      601 NRVKGEKRIPLVRLPYMVEHTVEKNGNLIIPHKHVHNIKFAWFDHDTYKAPNGYTTLED 660
Db      621 NRVKGEKRIPLVRLPYMVEHTVEKNGNLIIPHKHVHNIKFAWFDHDTYKAPNGYTTLED 680
QY      661 LFATIKYVVEHPDHPHSDNGWGNASEHVLGKDKHSDPNKNFKADPEPVEETPAEPEVP 720
Db      681 LFATIKYVVEHPDHPHSDNGWGNASEHVLGKDKHSDPNKNFKADPEPVEETPAEPEVP 740
QY      721 QVETEKVEAQLKEAEVLLAKVTSSSKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 780
Db      741 QVETEKVEAQLKEAEVLLAKVTSSSKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 800
QY      781 LLKGSNPSSVSKEKIN 796
Db      801 LLKGSNPSSVSKEKIN 816

```

; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10/387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 8
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-387-783-8

Query Match 87.3%; Score 695; DB 15; Length 819;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKENNRVSYIDGKATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 60
Db 21 SYELGLYQARTVKENNRVSYIDGKATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 80
QY 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKDKEDIVNEVGKGYVIVKDGKYYVYLKDA 120
Db 81 SHGDHYHYNGKVPYDAIISELLMKDPNYKDKEDIVNEVGKGYVIVKDGKYYVYLKDA 140
QY 121 AHADNVRKEEINRQKHSQHREGGTPRNDGAVALARSQRYTTDDGVIFNASDIIBDT 180
Db 141 AHADNVRKEEINRQKHSQHREGGTPRNDGAVALARSQRYTTDDGVIFNASDIIBDT 200
QY 181 GDAYIVPHGDHVIYIPKNELSASELAFAFUSGRGNLSNRTYRQNSDNTSRNNWVPS 240
Db 201 GDAYIVPHGDHVIYIPKNELSASELAFAFUSGRGNLSNRTYRQNSDNTSRNNWVPS 260
QY 241 VSNPGTTNTNNTSNTSQAQNSNDIDSLKQLYKPLSQRHVSVDGLVPDPAQITSR 300
Db 261 VSNPGTTNTNNTSNTSQAQNSNDIDSLKQLYKPLSQRHVSVDGLVPDPAQITSR 320
QY 301 ARGVAVPHGDHVIYIPYQMSLEBRARIIPLYRSNHWVPSRPEQSPQTPPEPSG 360
Db 321 ARGVAVPHGDHVIYIPYQMSLEBRARIIPLYRSNHWVPSRPEQSPQTPPEPSG 380
QY 361 POPAPNLKIDNSLSVQLVRKGVGYFEEKGISRYVFAKDLPSSETVKNLESKLSKQES 420
Db 381 POPAPNLKIDNSLSVQLVRKGVGYFEEKGISRYVFAKDLPSSETVKNLESKLSKQES 440
QY 421 VSHTLTAKKENVAPRDQFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
Db 441 VSHTLTAKKENVAPRDQFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 500
QY 481 KEKLVDDLLAFAPITHEPRLGKPNQSYTEDEVRIAQLADKYTTSDGYIFDEHDIISD 540
Db 501 KEKLVDDLLAFAPITHEPRLGKPNQSYTEDEVRIAQLADKYTTSDGYIFDEHDIISD 560
QY 541 EGDAYVTVMGHSHWIKGDSLSDEKVAQAAYTEKKGILPPSPDADVKANPTGDSAAAIY 600
Db 561 EGDAYVTVMGHSHWIKGDSLSDEKVAQAAYTEKKGILPPSPDADVKANPTGDSAAAIY 620
QY 601 NRVGKESRIPVRLPYMVEHTVEVKNGLIIPHKDHYHNIKPAWPDHHTYKAPNGYTTLED 660
Db 621 NRVGKESRIPVRLPYMVEHTVEVKNGLIIPHKDHYHNIKPAWPDHHTYKAPNGYTTLED 680
QY 661 LPATIKYVVEHDPHSDNGCNASEHVLGKHSEDPNKNFKADEPVEETPAEPVP 720
Db 681 LPATIKYVVEHDPHSDNGCNASEHVLGKHSEDPNKNFKADEPVEETPAEPVP 740
QY 721 QVETEKVBAQKEAEVLLAKVTDSSILKANATETLAGLNNLTLOIMDNNSIMAEAEKLLA 780
Db 741 QVETEKVBAQKEAEVLLAKVTDSSILKANATETLAGLNNLTLOIMDNNSIMAEAEKLLA 800

QY 781 LLKGSNPSSVSKEKIN 796
Db 801 LLKGSNPSSVSKEKIN 816

RESULT 5
US-10-324-143-20
; Sequence 20, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEE
; APPLICANT: CHARLEND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-20

Query Match 10.1%; Score 80; DB 15; Length 334;
Best Local Similarity 100.0%; Pred. No. 3.9e-70;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TVKENNRVSYIDGKATQKTENLTPEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 70
Db 12 TVKENNRVSYIDGKATQKTENLTPEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 71
QY 71 GKVPYDAIISELLMKDPNY 90
Db 72 GKVPYDAIISELLMKDPNY 91

RESULT 6
US-10-324-143-38
; Sequence 38, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEE
; APPLICANT: CHARLEND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38
; LENGTH: 690
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-38

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 840
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-7

Query Match      10.1%; Score 80; DB 15; Length 690;
Best Local Similarity 100.0%; Pred. No. 7.8e-70; Indels 0; Gaps 0;
Matches 80; Conservative 0; Mismatches 0;

QY 11 TVKENNRVSYIDGKQATKTKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHN 70
Db 12 TVKENNRVSYIDGKQATKTKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHN 71

QY 71 GKVPYDAIISEELMKDPNY 90
Db 72 GKVPYDAIISEELMKDPNY 91

RESULT 7
US-10-324-143-19
; Sequence 19, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEF
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELLETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 821
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-19

Query Match      10.1%; Score 80; DB 15; Length 821;
Best Local Similarity 100.0%; Pred. No. 9.1e-70; Indels 0; Gaps 0;
Matches 80; Conservative 0; Mismatches 0;

QY 11 TVKENNRVSYIDGKQATKTKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHN 70
Db 12 TVKENNRVSYIDGKQATKTKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHN 71

QY 71 GKVPYDAIISEELMKDPNY 90
Db 72 GKVPYDAIISEELMKDPNY 91

RESULT 8
US-09-884-465A-7
; Sequence 7, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josef
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
```

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 840
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-7

Query Match      10.1%; Score 80; DB 10; Length 840;
Best Local Similarity 100.0%; Pred. No. 9.3e-70; Indels 0; Gaps 0;
Matches 80; Conservative 0; Mismatches 0;

QY 11 TVKENNRVSYIDGKQATKTKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHN 70
Db 31 TVKENNRVSYIDGKQATKTKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHN 90

QY 71 GKVPYDAIISEELMKDPNY 90
Db 91 GKVPYDAIISEELMKDPNY 110

RESULT 9
US-10-324-143-8
; Sequence 8, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEF
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELLETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 840
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-8

Query Match      10.1%; Score 80; DB 15; Length 840;
Best Local Similarity 100.0%; Pred. No. 9.3e-70; Indels 0; Gaps 0;
Matches 80; Conservative 0; Mismatches 0;

QY 11 TVKENNRVSYIDGKQATKTKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHN 70
Db 31 TVKENNRVSYIDGKQATKTKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHN 90

QY 71 GKVPYDAIISEELMKDPNY 90
Db 91 GKVPYDAIISEELMKDPNY 110

RESULT 10
US-10-412-862-10
; Sequence 10, Application US/10412862
; Publication No. US20040052781A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-685
; CURRENT APPLICATION NUMBER: US/10/412,862
```

```
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 10
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-412-862-10

Query Match          7.5%; Score 60; DB 12; Length 819;
Best Local Similarity 100.0%; Pred. No. 6e-50;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYNGKVPYDAIISELLMKDPNY 90
    |||||||
DB 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYNGKVPYDAIISELLMKDPNY 111
    |||||||

RESULT 11
US-10-324-143-32
; Sequence 32, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEF
; APPLICANT: CHARLEAU, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-32

Query Match          7.5%; Score 60; DB 15; Length 819;
Best Local Similarity 100.0%; Pred. No. 6e-50;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYNGKVPYDAIISELLMKDPNY 90
    |||||||
DB 33 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYNGKVPYDAIISELLMKDPNY 92
    |||||||

RESULT 12
US-10-412-850-10
; Sequence 10, Application US/10412850
; Publication No. US20040001836A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-686
; CURRENT APPLICATION NUMBER: US/10/412,850
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
```

```
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 10
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-412-850-10

Query Match          7.5%; Score 60; DB 15; Length 819;
Best Local Similarity 100.0%; Pred. No. 6e-50;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYNGKVPYDAIISELLMKDPNY 90
    |||||||
DB 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYNGKVPYDAIISELLMKDPNY 111
    |||||||

RESULT 13
US-10-387-783-10
; Sequence 10, Application US/10387783
; Publication No. US20040005331A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10/387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 10
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-387-783-10

Query Match          7.5%; Score 60; DB 15; Length 819;
Best Local Similarity 100.0%; Pred. No. 6e-50;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYNGKVPYDAIISELLMKDPNY 90
    |||||||
DB 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYNGKVPYDAIISELLMKDPNY 111
    |||||||

RESULT 14
US-09-769-787-194
; Sequence 194, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P211290
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 194
```

```
; LENGTH: 826
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-194

Query Match      7.5%  Score 60; DB 10; Length 826;
Best Local Similarity 100.0%; Pred. No. 6e-50;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 90
Db 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 111

RESULT 15
US-09-884-465A-8
; Sequence 8, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-8

Query Match      7.5%  Score 60; DB 10; Length 838;
Best Local Similarity 100.0%; Pred. No. 6.1e-50;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 90
Db 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 111

RESULT 16
US-10-412-862-4
; Sequence 4, Application US/10412862
; Publication No. US20040052781A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-685
; CURRENT APPLICATION NUMBER: US/10/412,862
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-412-862-4
```

```
Query Match      7.5%  Score 60; DB 12; Length 838;
Best Local Similarity 100.0%; Pred. No. 6.1e-50;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 90
Db 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 111

RESULT 17
US-10-324-143-9
; Sequence 9, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: Hamel, Josee
; APPLICANT: Charland, Nathalie
; APPLICANT: Brodeur, Bernard R.
; APPLICANT: Martin, Denis
; APPLICANT: Blais, Normand
; APPLICANT: Ouellette, Catherine
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-9

Query Match      7.5%  Score 60; DB 15; Length 838;
Best Local Similarity 100.0%; Pred. No. 6.1e-50;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 90
Db 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 111

RESULT 18
US-10-412-850-4
; Sequence 4, Application US/10412850
; Publication No. US20040001836A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-686
; CURRENT APPLICATION NUMBER: US/10/412,850
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-412-850-4

Query Match      7.5%  Score 60; DB 15; Length 838;
Best Local Similarity 100.0%; Pred. No. 6.1e-50;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 90
Db 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 111
```

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPEVSKREGINAEQIVIKITDQGYVTSBGDHYHYNGKVPYDAIISBELLMKDPNY 90
|||||
DB 52 ENLTPEVSKREGINAEQIVIKITDQGYVTSBGDHYHYNGKVPYDAIISBELLMKDPNY 111
|||||

RESULT 19

US-10-387-783-4
; Sequence 4, Application US/10387783
; Publication No. US20040005331A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John B.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10/387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-387-783-4

Query Match 7.5%; Score 60; DB 15; Length 838;

Best Local Similarity 100.0%; Pred. No. 6.1e-50; Length 838;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPEVSKREGINAEQIVIKITDQGYVTSBGDHYHYNGKVPYDAIISBELLMKDPNY 90
|||||
DB 52 ENLTPEVSKREGINAEQIVIKITDQGYVTSBGDHYHYNGKVPYDAIISBELLMKDPNY 111
|||||

RESULT 20

US-10-324-143-22
; Sequence 22, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEF
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELLETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 613
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-22

Query Match 7.0%; Score 56; DB 15; Length 613;
Best Local Similarity 100.0%; Pred. No. 4.2e-46;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 RTNWVPSVSNFGTNTNTSNTNSQASQSNDDSLKQLYKLPQRSQHVESDGL 289
|||||
DB 27 RTNWVPSVSNFGTNTNTSNTNSQASQSNDDSLKQLYKLPQRSQHVESDGL 82
|||||

RESULT 21

US-09-765-272-66
; Sequence 66, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 66:

Query Match 6.4%; Score 51; DB 9; Length 763;
Best Local Similarity 100.0%; Pred. No. 4.6e-41;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 QGRVTTDDGYTFNADSIIEDTGDYIVPHGDHYHYIPKXNLSASELAAAAA 210
|||||
DB 159 QGRVTTDDGYTFNADSIIEDTGDYIVPHGDHYHYIPKXNLSASELAAAAA 209
|||||

RESULT 22

US-09-884-465A-12
; Sequence 12, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josef
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20

```

;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;
; COMPUTER: HP Vectra 486/33
;
; OPERATING SYSTEM: MSDOS version 6.2
;
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/09/765,272
;
; FILING DATE: 22-Jan-2001
;
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER: 08/961,083
;
; FILING DATE: <Unknown>
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: Brookes, A. Anders
;
; REGISTRATION NUMBER: 36,373
;
; REFERENCE/DOCKET NUMBER: PB340P2
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (301) 309-8504
;
; TELEFAX: (301) 309-8512
;
; INFORMATION FOR SEQ ID NO: 182:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 447 amino acids
;
; TYPE: amino acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 182:
;
US-09-765-272-182

Query Match          4.0%; Score 32; DB 9; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.9e-22;
Matches 32; Conservative 0; Mismatches 0; Indels

QY 46 AEQIVIKITDQGYVTSHGHDHYHNGKVPYDA 77
Db 43 AEQIVIKITDQGYVTSHGHDHYHNGKVPYDA 74

RESULT 25
US-09-769-787-38
; Sequence 38, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-38

Query Match          4.0%; Score 32; DB 10; Length 484;
Best Local Similarity 100.0%; Pred. No. 2e-22;
Matches 32; Conservative 0; Mismatches 0; Indels

QY 46 AEQIVIKITDQGYVTSHGHDHYHNGKVPYDA 77
Db 67 AEQIVIKITDQGYVTSHGHDHYHNGKVPYDA 98

Search completed: October 1, 2004, 07:41:06
Job time : 92 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 1, 2004, 07:35:26 ; Search time 7905 Seconds
(without alignments)
4364.460 Million cell updates/sec

Title: US-09-765-271-56
Perfect score: 4165
Sequence: 1 SYELGLYQARTVKENNRVSY.....KLALLKGSNPFSSVSKKIN 796

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 90 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xih
-Q=/cgn2.1/USPTO.spool_h/US09765271/runat_30092004.113753.12277/app_query.fasta_1.967
-DB=GenEmbl -QWRT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=90
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=25 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09765271.qcgn 1.1 5239 @runat_30092004.113753.12277 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	4163	100.0	2389	6	AR120265 Sequence
2	4163	100.0	2389	6	AR340956 Sequence
3	4163	100.0	2389	6	BD063274 Streptoco
4	4163	100.0	2406	6	AX569139 Sequence
5	4163	100.0	2451	6	BD268048 Streptoco
6	4163	100.0	2451	6	AR344447 Sequence
7	4163	100.0	2541	1	AF291695 Sequence
8	4163	100.0	8195	6	AR218862 Sequence
9	4163	100.0	8195	6	BD003774 Polynucle
c 10	4163	100.0	10256	1	AE007418 Streptoco
c 11	4163	100.0	349980	6	AX571763 Sequence
c 12	4159	99.9	10320	1	AE008479 Sequence
13	4112	97.7	20035	2	SPNEU1915
14	3236	77.7	2535	1	AF340221
15	3218	77.3	2523	6	BD263585
16	3218	77.3	2523	6	AX343072 Sequence
17	3218	77.3	2647	6	BD263590
18	3218	77.3	2647	6	AX343073 Sequence
19	2812	67.5	2478	6	BD229972 Human com
20	2797	67.2	2457	6	AX569137 Sequence
21	2795	67.1	2481	6	BD264091 Streptoco
22	2786	66.9	2457	1	AF318954 Streptoco
23	2786	66.9	2531	6	BD268049
24	2786	66.9	2531	6	AR344448 Sequence
25	2781	66.8	232807	2	SPNEU1901
26	2772	66.6	2517	1	AF318955
27	2772	66.6	2517	6	AX568778
28	2772	66.6	2531	6	BD268046
29	2772	66.6	2531	6	AR344445
30	2772	66.6	11931	1	AE007403
31	2772	66.6	349980	6	AX571762 Sequence
32	2750	66.0	2639	6	BD263591
33	2750	66.0	2639	6	AX343074
34	2733	65.6	12372	1	AE008464
35	2649	63.6	2290	6	AR120270
36	2649	63.6	2290	6	AR340961
37	2649	63.6	2290	6	BD063279
38	2431	58.4	2166	12	AF340222
c 39	2382	57.2	75874	2	SPNEU1907
40	1774	42.6	2359	6	AR219011
41	1774	42.6	2359	6	BD003923
42	1591	38.2	3171	6	BD263631 Polynucle
43	1298	31.2	75248	6	AX602204
c 44	1298	31.2	98050	1	SAG766854
45	1247	29.9	3120	6	BD263584
46	1247	29.9	3120	6	AX343070
47	1247	29.9	5048	6	BD263589
48	1247	29.9	5048	6	AX343071
49	1246	29.9	3117	6	AX568780
50	1246	29.9	3120	1	AF318956

```

c 51 1222.5 29.4 6867 6 AR218960 Sequence
c 52 1222.5 29.4 6867 6 BD003872 Polynucle
c 53 1216.5 29.2 21397 1 AE014279 Streptoco
c 54 1204 28.9 702 2 SPNEU1929 Streptoco
55 1203.5 28.9 1455 6 BD263587 Streptoco
56 1203.5 28.9 1455 6 BD264131 Streptoco
57 1203.5 28.9 1455 6 BD268047 Streptoco
58 1203.5 28.9 1455 6 AR344446 Sequence
59 1203.5 28.9 1455 6 BD223895 Nucleic a
60 1187.5 28.5 1342 6 AR120328 Sequence
61 1187.5 28.5 1342 6 AR341019 Sequence
62 1187.5 28.5 1342 6 BD063337 Streptoco
63 995 23.9 1446 6 AX608407 Sequence
64 988 23.7 973 6 AR219123 Sequence
65 988 23.7 973 6 BD004035 Polynucle
66 966.5 23.2 2528 6 BD263592 Novel str
67 966.5 23.2 2528 6 AX343078 Sequence
68 950 22.8 2478 6 AX088372 Sequence
69 950 22.8 10825 1 AE006623 Streptoco
70 947 22.7 10618 1 AE010110 Streptoco
71 946 22.7 2469 6 BD263632 Novel str
72 946 22.7 2469 6 AX607520 Sequence
73 946 22.7 5215 6 BD136831 Group B s
74 946 22.7 11617 6 AX602163 Sequence
75 946 22.7 199050 1 SAG766850 Streptoco
76 945 22.7 2478 1 AB073859 Streptoco
77 941 22.6 2472 6 BD263633 Novel str
78 941 22.6 50354 1 AE014169 Streptoco
79 941 22.6 323825 1 AP005146 Streptoco
80 937 22.5 21414 1 AE014248 Streptoco
81 934 22.4 2469 6 AX088376 Sequence
82 934 22.4 3501 1 AF062533 Streptoco
83 929 22.3 83 2469 6 BD263929 Nucleic a
84 841 20.2 1146 6 BD263932 Nucleic a
85 833 20.0 10312 1 AE007404 Streptoco
86 770.5 18.5 1684 6 AR219026 Sequence
87 770.5 18.5 1684 6 BD003938 Polynucle
88 679.5 16.3 1910 1 SAG290952 Streptoco
89 650 15.6 2721 6 AX799438 Sequence
90 644 15.5 816 6 AX799436 Sequence

```

ALIGNMENTS

```

RESULT 1
LOCUS AR120265 2389 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 55 from patent US 6159469.
ACCESSION AR120265
VERSION AR120265.1 GI:14103841
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2389)
AUTHORS Choi,G.H., Kunsch,C.A., Barash,S.C., Dillon,P.J., Dougherty,B.,
Fannon,W.R. and Rosen,C.A.
TITLE Streptococcus pneumoniae antigens and vaccines
JOURNAL Patent: US 6159469-A 55 12-DEC-2000;
FEATURES
Location/Qualifiers
1..2389
/organism="unknown"
/moi_type="unassigned DNA"
source

```

ORIGIN

```

Alignment Scores:
Pred. No.: 7.42e-230 Length: 2389
Score: 4163.00 Matches: 796
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.95% Indels: 0
DB: 6 Gaps: 0

```

```

US-09-765-271-56 (1-796) x AR120265 (1-2389)
QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnArgValSerTyr 20
Db 2 TCTTACGAGTTGGGAGCTGTATCAAGCTAGAAGCGTTAAGGAAAATPATCGTGTTCCTAT 61
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 62 ATAGATGAAACCAAGCCGCAAAACGAGAAATTTGACTCTCTGATGAGGTAGCAAG 121
QY 41 ArgGluGlyLysAsnAlaGluGlnLleValLleLysLleThrAspGlnGlyTyrValThr 60
Db 122 CGTGAAGGAATCAATCTGCTGAGCAATCTCATCAAGATAACAGACCAAGGCTATGCTACT 181
QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db 182 TCACATGGCGGACCATCTATCATTTATCAATGTTAGGTTCTTATGACGCTATCATCAGT 241
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspLleValAsnGlu 100
Db 242 GAAGATTTACTCATCAAGAGATCCAAACTATAAGCTAAAGATGAGGATATTGTTAATGAG 301
QY 101 ValLysGlyLysTyrValLleLysValAspGlyLysTyrTyrValTyrLleLysAspAla 120
Db 302 GTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGCT 361
QY 121 AlaHisAlaAspAsnValArgThrLysGluGluLleAsnArgGlnLysGlnGluHisSer 140
Db 362 GCCCAGCGGGAATGCTCGTCAAAAGAGGAAATCAATCGCAAAAACCAAGGCATAGT 421
QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
Db 422 CAACATCGTGAAGTGGAACTCCAGAAACGATGGTGTGTTGCTTGGCCTTGGCAAGTTCGCAA 481
QY 161 GlyArgTyrThrThrAspAspGlyTyrLlePheAsnAlaSerAspLleIleGluAspThr 180
Db 482 GGACGCTATCTACACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGATCT 541
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db 542 GGTGATGCTTATATCGTTCCTCATGGAGATCATACCATTACATTACCTTAAGAATGAGTTA 601
QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db 602 TCAGCTAGCGAGTTGGCTGCTGCAGAAAGCTTCTCTATCTGTGTCGAGGAAATCTGTCAAAT 661
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
Db 662 TCAAGAACCTATCGCCGCAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTCT 721
QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db 722 GTAAGCAATCCAGGAACCTACAATACTAACACAGCAACACAGCAACACTAACAGTCAA 781
QY 261 AlaSerGlnSerAsnAspLleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
Db 782 GCAAGTCAAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAAACTGCGCTTGAAT 841
QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnLleThrSerArgThr 300
Db 842 CAACGACATGTGAATCTGATGGCCTTGCTTTTGATCCAGCACAATATCAAGTCCGACACA 901
QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db 902 GCTAGAGGTGTTCAGTGGCCACACGAGATCATACCACCTTCATCCTTACTCTCAAAATG 961
QY 321 SerGluLeuGluGluArgLleAlaArgLleIleProLeuArgTyrArgSerAsnHisTrp 340
Db 962 TCTGAATTTGAGAGAGCAATCGCTCGCTATTAATTTCCCTTCTGTTATCGTTCAAACTATGG 1021
QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db 1022 GTACCAAGATTCAAGGCCGAGNACCAACCAAGTCCACACCGACTCCGGNACCTAGTCCAGGC 1081

```

QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerClnleuVal 380
Db 1082 CGCAACCTCGCACCAATCTTAAATAGACTCAAAATCTCTTTGGTAGTCAGCTGGTA 1141
QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
Db 1142 CGAAAAGTTGGGAAGGATATGTAATTCGAAGAAAAGGCACTCTCGTATGCTCTTGGG 1201
QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
Db 1202 AAAGATTTTACCATCTCGAAATCTGTTAAATCTTGAAGCAAGTTATCAAAAACAGAGAGT 1261
QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
Db 1262 GTTTCACACACTTAACTGCTTAAAGAAAATGTTGCTCTCGTGACCAAGAAATTTAT 1321
QY 441 AspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db 1322 GATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCTTGTGTTGNAATAAAGGTCGT 1381
QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuGluArgLeuAsnAspGluSerThrAsn 480
Db 1382 AATCTGATTTCCAGGCTTAGACAAATTTATAGAACGCTTGAATGATGAATCGACTAAT 1441
QY 481 LysGluLysLeuValAspAspLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db 1442 AAAGAAAATTTGTAGATGATTTATGTCATCTTAGCACCANTTACCATCCAGAGGGA 1501
QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
Db 1502 CTTGGCAAAACCAATCTCAAAATGAGTATATCTGAAGACGAAGTTGCTGATTTGCTCAATTA 1561
QY 521 AlaAspLysTyrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
Db 1562 GCTGATAGTATACAAGCTCAGATGGTTACATTTTGTATGAACATGATATATATCAGTGAT 1621
QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560
Db 1622 GAAGGAGATGATATGTAACGCTCATATGGCCATAGTCTAGTGGATTTGAAAAGATAGC 1681
QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
Db 1682 CTTTCTGATAGGAAAAGTTGACGCTCAAGCTTATCTAAGAAAAGGTTATCTACCT 1741
QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600
Db 1742 CCATCTCCAGACGAGATGTTTAAAGCAAAATCCAACTGGAGATAGTCAGCAGCTATTTC 1801
QY 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
Db 1802 AATCGTGTGAAGGGGAAAACGAATTCCTCGTTCGACTTCATATATGTTGAGCAT 1861
QY 621 ThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIle 640
Db 1862 ACAGTTGAGGTTAAACCGGTAATTTGATTTATCTCATAGGATCATTAACATAATATT 1921
QY 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660
Db 1922 AAATTTGCTGTTTGAATGATCACACATACAAAGCTCCAAATGCTATACCTTGGAAAGAT 1981
QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
Db 1982 TTGTTTGGCAGATTAAGTACTAGTAGAACCCCTGACGAACCTCCACATCTTAATGAT 2041
QY 681 GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
Db 2042 GGATGGGCAATGCGAGTGTGTAGGCAAGAAAGACACAGTGAAGATCCAAAT 2101
QY 701 LysAsnPheLysAlaAspGluProValGluGluThrProAlaGluProGluValPro 720
Db 2102 AAGAACTTCAAGCGGATGAAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCCT 2161

QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
Db 2162 CAAGTAGAGACTGAAAAAGTAGAACCCCACTCAAAAGAGCAGAAAGTTTGTCTCGGAA 2221
QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
Db 2222 GTAAAGGATTTAGTCTGAAAGCCATGCAACAGAACTCTAGCTGGTTTACGAATTAAT 2281
QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuAla 780
Db 2282 TTGACTCTTCAATATATGATAAACAATAGTATCATGCGAGAGCAGAAAAATTTACTGCG 2341
QY 781 LeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796
Db 2342 TTGTTAAAGGAAGTATCTCTTCATCTGTAAGTAAGGAAAAATAAAC 2389
RESULT 2
AR340956 2389 bp DNA linear PAT 17-AUG-2003
LOCUS AR340956 Sequence 55 from patent US 6573082.
DEFINITION AR340956
ACCESSION AR340956
VERSION AR340956.1 GI:33732935
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2389)
AUTHORS Choi, G.H., Kunsch, C.A., Barash, S.C., Dillon, P.J., Dougherty, B.,
Fannon, M.R. and Rosen, C.A.
TITLE Streptococcus pneumoniae antigens and vaccines
JOURNAL Patent: US 6573082-A 55 03-JUN-2003;
FEATURES Location/Qualifiers
source 1..2389
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 7.42e-230 Length: 2389
Score: 4163.00 Matches: 796
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.95% Indels: 0
DB: 6 Gaps: 0
US-09-765-271-56 (1-796) x AR340956 (1-2389)
QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db 2 TCTTACGATTTGGAGCTGTATCAAGCTAGAACGGTTAAGGAAAATATCTGTTTCTCAT 61
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 62 ATAGATGAAAAACAAGCGACGCAAAAAACGAGAAATTTGACTCTGATGAGTTAGCAAG 121
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db 122 CGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCACT 181
QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db 182 TCATGCGGACCACTATCATATTATACATGGTAAGGTTCTCTTATGACGCTATCATCAGT 241
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db 242 GAAGAATTTACTCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATTTGTAATGAG 301
QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db 302 GTCAAGGGTGGATGTATGTTATCAAGGTAGTAGGAAAAATCTATGTTTACCTTAAGGATGCT 361
QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140

Db 362 GCCACCGGATACCGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAGT 421
QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
Db 422 CAACATCGTGAAGGTGGAACCTCAAGAAACGATGGTGTCTGGCTTGGCAGCTTCGCAA 481
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db 482 GGACCTATACATACAGATGATGGTATATCTTTAATGCTTCGTGATATCATAGAGATACT 541
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db 542 GGTGATGCTTATATGCTTCTCATGGAGATCATTAACATTACCTTAAGAAATGAGTTA 601
QY 201 SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db 602 TCAGTAGCAGGTGGCTGCTCAGAGACCTTCTATCTGGTCGAGGAAATCTGCAAAAT 661
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
Db 662 TCAAGAACCTATCGCGACAAATAGCGTAAACATCTCAAGAACAAACCTGGGTACCTTCT 721
QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db 722 GTAAGCAATCCAGGAATCAAAATCTAAACACAAAGCAACACAGCAACAACTAACAGTCAA 781
QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrIleLeuProLeuSer 280
Db 782 GCAAGTCAAAAGTAATGACATTCATGATGCTCTTGAACACAGCTCTACAAATGGCTTTGAGT 841
QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db 842 CAACGACATGTAGATCTGATGGCTTGTCTTTGATCCAGACCAATCACAAGTCGACAA 901
QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db 902 GCTAGAGTGTTCAGTGCACACGAGATCATTAACCATTCATCCCTTACTCTCAATG 961
QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340
Db 962 TCTGAATTCGGAAGACGAATCGCTGATATATTCCTTCCTGTTATCGTTTCAACCAATGG 1021
QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db 1022 GTACAGATTCAGGCGCAGACCAACCAAGTCCACACCGACCTCGGAACTAGTCAGGC 1081
QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
Db 1082 CGGCAACCTGCACCAATCTTAAATATAGACTCAAAATTCCTTTGTTAGTCAGCTGGTA 1141
QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
Db 1142 CGAAAAGTTGGGAAGGATGATGTTTCGAAGAAAGGGATCTCTGTTATGTTCTTGGC 1201
QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
Db 1202 AAGATTTACCATCTGAACTGTAAATCTTTGAAGCAAGTTATCAAAACAAGAGAT 1261
QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
Db 1262 GTTTCACACACTTAACTGCTTAAAGAAAGAAATGTTGCTCTCGTCAACCAAGAAATTTAT 1321
QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db 1322 GATAAGCATATATCTGTTAATCTGAGGCTCATTAAGCCTTGTGTTGAAATAAGGTCGT 1381
QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
Db 1382 AATTCCTGATTCGAAGCCTTAGACAAATATTAGAACGCTTGAATGAATCAATCACTAAT 1441
QY 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db 1442 AAAGAAAAATTTGGTAGATGATTTATTGGCATTCCTAGCACCAATTAACCATCCAGAGCGA 1501

QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
Db 1502 CTTGGCAACCAATCTCAATTTGAGTATCTGAAGCAAGTTCGTTATTGCTCAATTA 1561
QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
Db 1562 GCTGATAAGTATACAACTGATGTTTACATTTTGTATGAAATGATATATATCATAGTAT 1621
QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560
Db 1622 GAAGAGATGATATGTAACGCTCATATGGCCCATAGTCACTGGATTGGAAAAGATAGC 1681
QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
Db 1682 CTTTCTGATAGGAAAAAGTTGCCAGTCAAGCCTATCTAAAGAAAAAGGTATCTACCT 1741
QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600
Db 1742 CCATCTCCAGACCGCAGATGTTAAAGCAAAATCCAACTGGAGATAGTCAGCAGCTATTTAC 1801
QY 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
Db 1802 AATCGTGTGAAGAGGAAAAACGAATTCACCTCGTTGCACTTCCATATATGTTGAGCAT 1861
QY 621 ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle 640
Db 1862 ACAGTTGAGTTAAAAACGGTAATTTGATTTATTCCTCATAGGATCATTAACATAATTT 1921
QY 641 LysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660
Db 1922 AAAATTTGCTTGGTTTGGATGATCACACATACAAAGCTCCAAATGGCTATATACCTTGAAGAT 1981
QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
Db 1982 TTGTTTGGCAGCATTAAGTACTACGTAGAACACCTTCAGCACTCCACATTTCTATGAT 2041
QY 681 GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
Db 2042 GGATGGGCAATGCCAGTGAGCATGTGTGTAGCAAGAAAGACCAAGTGAAGATCCAAAT 2101
QY 701 LysAsnPheLysAlaAspGluLeuProValGluGluThrProAlaGluProGluValPro 720
Db 2102 AAGAACTTCAAAAGCGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCCT 2161
QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
Db 2162 CAAGTAGAGACTGAAAAGTAGAGCCCAACTCAAGAAACAGAGTTTGTTCGCAAA 2221
QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
Db 2222 GTAACCGATTTCTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGTGTGTTTACGAAATAAT 2281
QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780
Db 2282 TTGACTCTTCAATTTATGATTAACATAGTATCATGCGAAGACAGAAAAATTTACTTGG 2341
QY 781 LeuLeuLysGlySerAsnProSerValSerLysGluLysIleAsn 796
Db 2342 TTGTTAAAGGAGTAAATCTTCACTGTAGTAAGGAAAAATAAATAAAC 2389

RESULT 3

BD063274
LOCUS 2389 bp DNA linear PAT 27-AUG-2002
DEFINITION Streptococcus pneumoniae antigens and vaccines.
ACCESSION BD063274
VERSION BD063274.1 GI:22608877
KEYWORDS JP 2001505415-A/28.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 2389)
AUTHORS Kunsch, C.A., Choi, G.H., Johnson, S.L. and Hromockyj, A.

TITLE
JOURNAL
Patent: JP 2001505415-A 28 24-APR-2001;
HUMAN GENOME SCIENCES INC
COMMENT
PN JP 2001505415-A/28
PD 24-APR-2001
PF 30-OCT-1997 JP 1998520667
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH, GIL H CHOI, SYDNOR L JOHNSON, ALEX HROMOCKY, J PC
C12N15/31, C12N5/18, C12N1/21, C07K14/315, C12Q1/68, A61K39/09, PC
G01N33/569
PC G01N33/68
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers.

FEATURES
source
1..2389 Location/Qualifiers.
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

ORIGIN
Alignment Scores:
Pred. No.: 7,42e-230 Length: 2389
Score: 4163.00 Matches: 796
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.95% Indels: 0
DB: 6 Gaps: 0

US-09-765-271-56 (1-796) x BD063274 (1-2389)

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnArgValSerTyr 20
Db 2 TCTTACGAGTTGGACTGTATCAAGCTAGACGGTTAGGAAATATTCGTTTCCTAT 61
Qy 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 62 ATAGATGAAACCAAGCGACGCAAAACCGAGAATTTGACTCTCTGATGAGGTTAGCAAG 121
Qy 41 ArgGluGlyIleAsnAlaGluGlnIleValLysIleThrAspGlnGlyTyrValThr 60
Db 122 CGTAAGGAATCAATGCTGACAAATCGTATCAAGATAACAGACCAAGGCTATGCTACT 181
Qy 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleSer 80
Db 182 TCACATGGCGACCACTATCATATTAACATGTAAGGTTCTTATGACGCTATCATCAGT 241
Qy 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db 242 GAAGAATTAATCAATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATTTGTTAATGAG 301
Qy 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db 302 GTCNAGGGTGATATGTTATCAAGTATGATGGAATAACTATGTTTACCTTTAAGGATGCT 361
Qy 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
Db 362 GCCCACGCGATACGTCGCTACAAAGAGGAATCAATCGACAAAACAGACCATAGT 421
Qy 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
Db 422 CAACATCGTGAAGGTGAATCCCAAGAAACCGATGGTGTGTTGCTGGCAGCGTTCCGAA 481
Qy 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db 482 GGAGCTTACTACTACAGATGATGGTTATATCTTTAATGCTCTGTATATCATAGAGATACT 541
Qy 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db 542 GGTGATCTTATATCGTTCTCATGGAGATCATTTACCATTTACATTTCTTAAGAATGAGTTA 601
Qy 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220

Db 602 TCAGCTAGCGAGTTGGCTGCTGCAGAACCTTCCTATCTGGTCGAGGAAATCTGTCAAAAT 661
Qy 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
Db 662 TCAGAACCTATCGCCGCAAAATAGCGATAACATTCAGAGAACAACTGGGTACCTTCT 721
Qy 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db 722 GTAACCAATCCAGGAACATACAAATACAAACAGCAACACAGCAACACTAACAGTCAA 781
Qy 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
Db 782 GCAAGTCAAAGTATGACATTTGATAGTCTCTTGAACAGCTCTCAAACTGCTTTGAGT 841
Qy 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db 842 CAACGACATGTAGAATCTGATGGCTTGTCTTTGATCGACACAAATCACAACTCGAACA 901
Qy 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db 902 GCTAGAGGTGTTCAGTGCACACGAGATCATTAACCACTTCATCCCTTACTCTCAAAATG 961
Qy 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340
Db 962 TCTGAATTTGGAAGAACGATCGCTGATATTTATCCCTTCTGTTATCGTTCAACACTTGG 1021
Qy 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db 1022 GTACCAGATTCAAGGCCAGAACCAACCAAGTCCCAACCGACTCCGGAACCTAGTCCAGC 1081
Qy 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
Db 1082 CGCAACCTCGCACCAATCTTAAATATAGACTCAAATTTCTTTGGTTAGTCAGCTGTA 1141
Qy 381 ArgLysValGlyGluGlyTyrValPheGluGlyLysGlyIleSerArgTyrValPheAla 400
Db 1142 CGAAAAGTTGGGGAAGGATATGTTATCGAAGAAAAGGCACTCTCTGTTATGTTTGGC 1201
Qy 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
Db 1202 AAAGATTATCCATCTGAACCTGTTTAAATACTTGAAGCAAGTTATCAAAACAGAGAGT 1261
Qy 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
Db 1262 GTTTCACACACTTTAACTGCTTAAAAAAGAAAATGCTCTCTCGTGACCAAGAAATTTAT 1321
Qy 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db 1322 GATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATTAAGGTCGT 1381
Qy 461 AsnSerAspPheGlnAlaLeuAspLysLeuGluArgLeuAsnAspGluSerThrAsn 480
Db 1382 AATTCTGATTTCAGAGCCTTAGACAAATTTATAGAACGCTTGAATGATGAATCGACTAAT 1441
Qy 481 LysGluLysLeuValAspAspLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db 1442 AAAGAAAATTTGGTAGATGATTTATTGGCATTTCTAGCACCAATTAACCATCCAGAGCGA 1501
Qy 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
Db 1502 CTTGGCAACCAAAATCTCAAAATGATATCTAGAACGCAAGTTCTGTTATGCTCAATTA 1561
Qy 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
Db 1562 GCTGATAAGTATACACAGTCAGATGGTTACATTTTGTATCAACATGATATAATCAGTGAT 1621
Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560
Db 1622 GAAGGAGATGCATATGTAAACGCTCATATGGGCGCATAGTCACCTGAGTTGAAAAGATAGC 1681
Qy 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
Db 1682 CTTTCTGATNAGAAAAGTTGACGCTCAAGCCTATATAAGAAAAGGATATCTTACCT 1741

QY	591	ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaAlaIleTyr	600	Query Match: 99.95%	Indels: 0
DB	1742	CCATCTCAGACGCGAGATGTTAAAGCAATCAATCGGAGATAGTCAGCAGCTAATTAC	1801	Gaps: 0	
QY	601	AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis	620	US-09-765-271-56 (1-796) x AX569139 (1-2406)	
DB	1802	AATCGTGTGAAGGGGAAACCAATTCACCTCGTTCGACTTCCATATATGTTGAGCAT	1861		
QY	621	ThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIle	640		
DB	1862	ACAGTTGAGTTAAAGCGTAATTTGATTATTCCTCATAGGATCATTAACCATATATT	1921		
QY	641	LysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp	660		
DB	1922	AAATTTGCTTGTGTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAGAT	1981		
QY	661	LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp	680		
DB	1982	TTGTTTGGACGATTAAAGTACTACGTAGAACACCCCTGACGAAGTCCACATTCATGAT	2041		
QY	681	GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn	700		
DB	2042	GGATGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCCACAGTGAAGATCCAAAT	2101		
QY	701	LysAsnPheLysAlaAspGluProValGluGluThrProAlaGluProGluValPro	720		
DB	2102	AAGAACCTTCAAGCGGATGAAGAGCCAGTAGAGGAACACCTGCTGAGCCAGAGTCCCT	2161		
QY	721	GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys	740		
DB	2162	CAAGTAGAGACTGAANAAGTAGAGCCCACTCAAGAAGACGAGAAGTTTGTTCGAAA	2221		
QY	741	ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrIleuAlaGlyLeuArgAsnAsn	760		
DB	2222	GTAACGATCTAGTCTGAAAGCCAATGCAACAGAACTCTAGCTGGTTTACGAAATAAT	2281		
QY	761	LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla	780		
DB	2282	TTGACTCTTCAAAATTATGATAACAATAGTATCATGGCAGACGACGAAANAATTACTTGG	2341		
QY	781	LeuLeuLysGlySerAsnProSerValSerLysGluLysIleAsn	796		
DB	2342	TTGTTAAAGGAAGTAATCTTTCATCTGTAAGTAAGGAAAAATAAAC	2389		
RESULT 4					
AX569139		2406 bp	DNA	linear	PAT 29-NOV-2002
LOCUS					
DEFINITION		Sequence 2347 from Patent WO02077021.			
ACCESSION		AX569139			
VERSION		AX569139.1			
KEYWORDS		GI:26002636			
SOURCE		Streptococcus pneumoniae			
ORGANISM		Streptococcus pneumoniae			
REFERENCE		Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
AUTHORS		Streptococcus.			
TITLE		1			
JOURNAL		Masnigani, V., Tettelin, H. and Fraser, C.			
FEATURES		Streptococcus pneumoniae proteins and nucleic acids			
source		Patent: WO 02077021-A 2347 03-OCT-2002;			
		Chiron Spa (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)			
		Location/Qualifiers			
		1. 2406			
		/organism="Streptococcus pneumoniae"			
		/mol_type="unassigned DNA"			
		/db_xref="taxon:1313"			
ORIGIN					
Alignment Scores:					
Pred. No.:		7.48e-230	Length:		2406
Score:		4163.00	Matches:		795
Percent Similarity:		99.87%	Conservative:		0
Best Local Similarity:		99.87%	Mismatches:		1

QY	1	SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr	20		
DB	19	TCTTACGAGTTGGGACTGTATCAAGCTAGAAGCGTTAAGGAAAAATAATCGTGTTCCTTAT	78		
QY	21	IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys	40		
DB	79	ATAGATGGAACCAAGCGCAACCAACCGGAGAAATTTGACTCTCATGATGAGGTAGCAAG	138		
QY	41	ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr	60		
DB	139	CGTAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACACAGCAAGGCTATGCTACT	198		
QY	61	SerHisGlyAspHisTyrTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer	80		
DB	199	TCACATGGCGACCACTATCATTAATTAATGTAAGGTTCCTTATGAGCTATCATCAGT	258		
QY	81	GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu	100		
DB	259	GAAGAATTACTCATGAGAGATCCAACTATAGCTTAAAGATGAGGATATTGTTAATGAG	318		
QY	101	ValLysGlyTyrTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla	120		
DB	319	GTCAGGGTGGATATGTTATCAAGGTAGATGAAAAATCTATGTTTACCTTAAGGATGCT	378		
QY	121	AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgLysGlnGluHisSer	140		
DB	379	GCCCAAGCGGATAAAGTCGTCGTAACAAAGAGAGAAATCAATCGACAAAAACAAGAGCATAGT	438		
QY	141	GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln	160		
DB	439	CAACATCGTGAAGGTGGAACTCCAGAAACGATGTGTGTTCCTTGGCACGTTTCGCAA	498		
QY	161	GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr	180		
DB	499	GGACGCTATACTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGATACT	558		
QY	181	GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu	200		
DB	559	GGTGTATGCTTATATCGTTCCTCATGGAGATCATTAACATTACATTCCTAAGAAATGAGTTA	618		
QY	201	SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn	220		
DB	619	TCAGCTAGCGAGTTGGCTGCTGCAGAGAGCTTCCATATCTGTCGAGGAAATCTGTCAAAT	678		
QY	221	SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer	240		
DB	679	TCAAGAACCTATCGCCGACAAATAAGCGATTAACACTTCAAGAACAACTGGGTACCTTCT	738		
QY	241	ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln	260		
DB	739	GTAAGCAATCCAGGAATCAAAATATACTAACACAGCAACAAACAGCAACACTACAGTCAA	798		
QY	261	AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer	280		
DB	799	GCAAGTCAAGTAATGACATTGATGATGCTCTTGAAACAGCTCTACAACTGCCTTTGAGT	858		
QY	281	GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr	300		
DB	859	CAACGACATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCACAAATCAACAGTCAACA	918		
QY	301	AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet	320		
DB	919	GCTAGAGGTGTCAGTGCACACGGAGATCATTAACATTCATCCCTTACTCTCAATG	978		
QY	321	SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp	340		
DB	979	TCTGAATTGAAGAACGAATCGCTCGTATTATTCCTTATTCCTTATTCCTTATTCCTTATTC	1038		

QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
 DB 1039 GTACCAAGTTCAGGCCAGAACACCAAGTCCACACCGACTCCGGACCTAGTCCAGGC 1098
 QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
 DB 1099 CCGCAACCTCGACCAATCTTAAATAGACTCAAAATCTCTCTTTGGTTAGTCAGCTGGTA 1158
 QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
 DB 1159 CGAAAGTTCGGGAAGGATATGATTTCGAAGAAAGGGCATCTCTCGTTATGCTTTGGC 1218
 QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
 DB 1219 AAGATTTACATCTGAACTGTAAATCTTGAAGCAAGTTATCAAAACAGAGAGT 1278
 QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
 DB 1279 GTTTTCACACTTTAACTGCTAAAGAAAGAAATGTTGCTCTCGTACCAAGAAATTTAT 1338
 QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
 DB 1339 GATAAGCATATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTAATAAAGGTCGT 1398
 QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuGluArgLeuAsnAspGluSerThrAsn 480
 DB 1399 AATCTGATTTCCAGCCTTAGACAAATATTAGAACGCTTGAATGATGACTAAT 1458
 QY 481 LysGluLysLeuValAspLeuLeuAlaPheLeuAlaProLleThrHisProGluArg 500
 DB 1459 AAAGAAAAATGTTAGATGATTTATTGGCATTTCTAGCACCAATATCCCATCCAGACGA 1518
 QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrLysAspGluValArgIleAlaGlnLeu 520
 DB 1519 CTTGGCAACCAAAATCTCAAAITGATGATGACTGAGACCAAGTCTGATTTGCTCAATTA 1578
 QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
 DB 1579 GCTGATAAGTATACACGCTCAGATGTTATCATTTTGTATGACATGATATATACAGTAT 1638
 QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560
 DB 1639 GAAGAGATGCATATGTAACGCTCATATGGCCATAGTCACTGGATTGGAAAGATAGC 1698
 QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
 DB 1699 CTTTCTGATAAGGAAAGTTGACGCTCAAGCCTATATAAGAAAGGATATCTTACCT 1758
 QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600
 DB 1759 CCATCTCCAGACGAGATGTTAAGCAATCCAACTGGAGATAGTCAGCAGCTATTATC 1818
 QY 601 AsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
 DB 1819 AATCGTGTGAAGGGGAAAAACCAATTCCTACTCGTTCGACTTCATATATGGTTGAGCAT 1878
 QY 621 ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle 640
 DB 1879 ACAGTTGAGGTTAAAAACCGTAATTTGATTATTTCTCATAGGATCATTAACCAATAATAT 1938
 QY 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660
 DB 1939 AAATTTCTGTTGTTGATGATCACATACAAAGCTCCAAATGGCTATATCTTTGGAGAT 1998
 QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
 DB 1999 TTGTTTGGCAGCATTAAGTACTAGTAGAACACCCCTGACCAACGCTCCCATCTTAATGAT 2058
 QY 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
 DB 2059 GSAATGGGGCAATCCCATGTAGCATGTGTTAGGCAAGAAAGACCACTGAGATCCAAAT 2118
 QY 701 LysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720

DB 2119 AAGAACTTCAAGCGGATGAGAGCCAGTAGAGAAACACCTGCTGAGCCAGAAAGTCCCT 2178
 QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuAlaLys 740
 DB 2179 CAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGCAGAAAGTTTGTCTTGCGAAA 2238
 QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
 DB 2239 GTAACGGATTCTAGTCTGAAAGGCCAATGCAACAGAACTCTAGCTGGTTTACGAAATAAT 2298
 QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuAla 780
 DB 2299 TTGACITCTTCAAAATATGATATCAATATGATATCATGCAAGCAGAGAAATTTACTTGGC 2358
 QY 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796
 DB 2359 TTGTTAAAGGAAGTAATCTTCTCATCTGTAAGTAAGGAAAAATAAATAC 2406
 RESULT 5
 BD268048
 LOCUS Streptococcus pneumoniae protein and immunogenic fragments for
 DEFINITION vaccines.
 ACCESSION BD268048
 VERSION BD268048.1 GI:33077816
 KEYWORDS JP 2002532561-A/6.
 SOURCE Streptococcus pneumoniae
 ORGANISM Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 REFERENCE 1 (bases 1 to 2451)
 AUTHORS Johnson, L.S., Koenig, S. and Adamou, J.E.
 TITLE Streptococcus pneumoniae protein and immunogenic fragments for
 JOURNAL Patent: JP 2002532561-A 6 02-OCT-2002;
 MEDIMMUNE INC
 COMMENT OS Streptococcus pneumoniae
 PN JP 2002532561-A/6
 PD 02-OCT-2002
 PF 21-DEC-1999 JP 2000589215
 PR 21-DEC-1998 US 60/113048
 PI LESLIE S JOHNSON, SCOTT KOENIG, JOHN E ADAMOU
 PC A61K39/09 A61K38/00 A61P31/10 A61K37/02
 CC Streptococcus pneumoniae protein and immunogenic fragments for
 CC vaccines
 FH Key Location/Qualifiers
 FT source 1..2451
 FT /organism='Streptococcus pneumoniae'.
 FEATURES
 source Location/Qualifiers
 1..2451
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /db_xref="taxon:1313"
 ORIGIN
 Alignment Scores:
 Pred. No.: 7,666-230 Length: 2451
 Score: 4163.00 Matches: 796
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.95% Indels: 0
 DB: 6 Gaps: 0
 US-09-765-271-56 (1-796) x BD268048 (1-2451)
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
 DB 61 TCTTACGAGTTGGAGCTGTATCAAGCTAGAACGGTTAAGGAAATATATCGTGTTCCTAT 120
 QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
 DB 121 ATAGATGGAAACCAACGCGACGCAAAAAACGGAATTTGACTCTCTGATGAGGTTAGCAAG 180

QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleIleValIleIleThrAspGlnGlyTyrValThr 60
DB 181 CGTGAAGAAATCAATCTGAGCAATCTCATCAGATAACAGACCAAGGCTATGCTACT 240
QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
DB 241 TCACATGCGCACCATATCATTTATTAATGTAAGGTTCTCTTATGACGCTATCATCAGT 300
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
DB 301 GAAGAATTTACTCATGAAAGATCCAACTATTAAGCTAAAGATGAGGATTTGTTAATGAG 360
QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysLysAspAla 120
DB 361 GTCAGGGTGGATGATGTTATCAAGGTAGATGGAAATACTATGTTTACTTAAAGATGCT 420
QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
DB 421 GCCCAGCGGATAACGTCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAGT 480
QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
DB 481 CAACATCGTAGAGGGGAACCTCCAGAAACGATGGTGTGCTTGTGCTTGGCAGGTTGCGCAA 540
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
DB 541 GGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATACT 600
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
DB 601 GGTGATGCTTATATGCTTCTCATGGAGATCATTAACCATTAATCTCTTAAGAAATGAGTTA 660
QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
DB 661 TCAGCTAGCAGTGGCTGCTCAAGAGCTTCTTATCTGTCGAGGAAATCTGCAAAAT 720
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
DB 721 TCAGAACCTATCGCGACAAATAGCGATAACACTTCAGAACAAACCTGGGTACTCTTCT 780
QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
DB 781 GTAAACCAATCCAGGAACCTACAAATACTAACACAGCAACAAACAGCAACACTAACAGTCAA 840
QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
DB 841 GCAAGTCAAGTAATGACATTTGATGCTCTTTGAAACAGCTCTCAAACTGCTCTTGAGT 900
QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
DB 901 CAACGACATGTAGAAATCTGATGCTTGTCTTTGATCCAGCACAAATCACAGTCCGAACA 960
QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
DB 961 GCTAGAGGTGGTCAGTGCACACGAGATCATTAACCACTTCATCCCTTACTCTCAAAATG 1020
QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyrP 340
DB 1021 TCTGAATTTGGAAGAGAAATCGCTCGTATTATTCCTTTCCTGTTATCTGTTTCAAAACCATTTGG 1080
QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
DB 1081 GTACCAAGATTTCAAGGCCAGAACCAACAGTCCACAAACGACTCCGGAACCTAGTCCAGGC 1140
QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
DB 1141 CGCAACCTGCAACCAATCTTAAATPAGACTCAAAATCTCTTCTTGTGTTAGTCAGCTGGTA 1200
QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
DB 1201 CGAAAAATTTGGGAAGGATATGTTATTCGAAGAAAGGGCATCTCTCGTTATGTCCTTGGC 1260

QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
DB 1261 AAAGATTTTACCATCTGAAACTGTTAAAAATCTTGAAGCAAGTTTATCAAAACAAGAGAGT 1320
QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
DB 1321 GTTTCACACACTTTAACTGCTAAAAAGAAAAATGTTGCTCTCGTCGACCAAGAAATTTAT 1380
QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
DB 1381 GATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCTTGTGTGNAATAAAGGTCGT 1440
QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
DB 1441 AATTCTGATTTCCAAAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAAT 1500
QY 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
DB 1501 AAAGAAAAATTTGGTAGATGATTTATTGGCATTTCTTAGCAACCAATTACCCTCCAGAGCGA 1560
QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
DB 1561 CTTGCGCAACCAATTTCTCAATTTAGTATACTGAGACGAGGTTCTGTTATGCTCAATTA 1620
QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
DB 1621 GCTGATAAGTATACAACTGATGTTTACATTTTTTGATGAACATGATATATCATGATGAT 1680
QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560
DB 1681 GAAGAGATGATATGTAACGCTCATATGGCCCATGTCATCTGATTTGGAAAAAGATAGC 1740
QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
DB 1741 CTTTCTGATAGAAAAAGTTGCGAGTCAAGCTATACTAAAGAAAAAGGTATCTTACTCT 1800
QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600
DB 1801 CCATCTCAGACGCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTCGACGAGTATTTAC 1860
QY 601 AsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
DB 1861 AATGCTGGAAGGGGAAAAACGAATTCACCTGCTTCCACTTCCATATATGTTGAGCAT 1920
QY 621 ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle 640
DB 1921 ACAGTTGAGTTAAAAACGGTAATTTGATTTATCTCATAGGATCATTAATATAT 1980
QY 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660
DB 1981 AAATTTGCTTGGTTTGTATGATCACACATACAAAGCTCCAAATGGCTATACCTTGAAGAT 2040
QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
DB 2041 TTGTTTGGCAGATTAACTACTTACGTAGAACACCTCGACGACGTCCTCAATTTCTAATGAT 2100
QY 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
DB 2101 GGATGGGCAATGCCAGTGAGCATGTTTAGGCAAGAAAGACCAACAGATGAAGATCCAAAT 2160
QY 701 LysAsnPheLysAlaAspGluProValGluGluThrProAlaGluProGluValPro 720
DB 2161 AAGAACTTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCAGAGTCCCT 2220
QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
DB 2221 CAAGTAGAGACTGAAAAAGTAGAACCCCACTCAAGAAAGCAAGAGTTTGTCTGCGAAA 2280
QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
DB 2281 GTACCGATTTCTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAAT 2340
QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780


```

Db      2341 TTGACTCTTCAAATATGATGATAACATAGTATCATGGCAGAGCAAGAAATTTACTTGGC 2400
Qy      781  LeuLeuLysGlySerAsnProSerValSerLysGluLysIleAsn 796
Db      2401 TTGTTAAAGGAAGTAATCTTCACTCTGTAAGTAAGGAAATAAAC 2448

RESULT 6
LOCUS      AR344447          2451 bp      DNA      linear      PAT 17-AUG-2003
DEFINITION Sequence 9 from patent US 6582706.
ACCESSION AR344447
VERSION   AR344447.1  GI:33740486
KEYWORDS
SOURCE    Unknown.
ORGANISM  Unknown.
REFERENCE 1 (bases 1 to 2451)
AUTHORS  Johnson, L.S. and Adamou, J.E.
TITLE    Vaccine compositions comprising Streptococcus pneumoniae
         polypeptides having selected structural MOTIFS
JOURNAL  Patent: US 6582706-A 9 24-JUN-2003;
FEATURES  Location/Qualifiers
           source          1..2451
                        /organism="unknown"
                        /mol_type="genomic DNA"

ORIGIN
Alignment Scores:  7.66e-230      Length: 2451
Pred. No.: 4163.00      Matches: 796
Score: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 99.95%      Indels: 0
DB: 6      Gaps: 0

US-09-765-271-56 (1-796) x AR344447 (1-2451)

Qy      1  SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnAtgValSerTyr 20
Db      61  TCTTACGAGTTGGACTGTATCAAGCTAGACGGTTAAGGAAATAATTCGGTTTCCTAT 120
Qy      21  IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db      121 ATAGATGGAACAAACGAGCAGCAAAACCGAGAATTTCCTCTGATGAGTTAGCAAG 180
Qy      41  ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db      181 CGTGAAGGATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCACT 240
Qy      61  SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db      241 TCACATGGCGACCACTATCATATTACATGGTAAAGTTCTTATGACGCTATCATCAGT 300
Qy      81  GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db      301 GAAGAATTACTCATGAAGATCCAAACTATAAGCTAAAGATGAGGATATTTGTTAATGAG 360
Qy      101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db      361 GTCNAGGGTGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGCT 420
Qy      121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
Db      421 GCCCAGCGGATAACGTCCTGCTACAAAGAGAAATCAATCGCAAAACAAAGAGCATAGT 480
Qy      141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyValAlaValAlaLeuAlaArgSerGln 160
Db      481 CAACATCGTGAAGGTGGAATCCCAAGAAACGATGGTGCTGTGCTTGGCCATCGCTCGCAA 540
Qy      161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db      541 GGACGCTACTACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATACT 600

```

```

Qy      181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db      601 CGTGTATGCTTATATGCTTCTCATGAGATCATTAACCATTAATCCTAAGAATGAGTTA 660
Qy      201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db      661 TCAGCTAGCGAGTTGGCTGCTCGAAGCCTTCCCTATCTGGTCGAGGAATCTGTCAAT 720
Qy      221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
Db      721 TCAGAAGCCTATCGCCGCAAAATAGCATACACTTCAAGAACAACTGGGTACTCTTCT 780
Qy      241 ValSerAsnProGlyThrThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db      781 GTAGCAATCCAGGAACCTACAAATATAACAACAGCAACAAACAGCAACACTAACAGTCAA 840
Qy      261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
Db      841 GCAAGTCAAGTATATGACATTTGATAGTCTCTTGAACACAGCTCTACAACTGCCTTTGAGT 900
Qy      281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db      901 CAACGACATGTAGAAATCTGATGGCCTTCTTTGATCCAGCACAAATCACAGTCCAAACA 960
Qy      301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db      961 GCTAGAGGTGTTCAGTGCACACGAGATCATTAACCACTTCACTCCCTTACTCTCAAAATG 1020
Qy      321 SerGluLeuGluGlyArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340
Db      1021 TCTGATTTGGAGAACGAAATCGCTCGTATTAATCCCTTCTGTTATCGTTCAAACCATTTGG 1080
Qy      341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db      1081 GTACGAGATTCAAGGCCAGAACCAACCAAGTCCACACGACTCCGGAACCTAGTCCAGGC 1140
Qy      361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
Db      1141 CCGCAACCTCGACCAAAATCTTAAATAGACTCAAATTTCTCTTTGGTTAGTCAGCTGTA 1200
Qy      381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
Db      1201 CGAAAGTTGGGAGAGGATATGATATTCGAAGAAAGGGCATCTCTCTGTTATGCTTTTCGG 1260
Qy      401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
Db      1261 AAGATTTACCATCTGAACCTGTTAAATAATCTTGAAAGCAAGTTATCAAAACAGAGAGT 1320
Qy      421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
Db      1321 GTTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCTCTGTCGACCAAGATTTTAT 1380
Qy      441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db      1381 GATAAGCATATTAATCTGTTAACTGAGGCTCAATAAGCCCTTGTGTGAAATAAGGGTCGT 1440
Qy      461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
Db      1441 AATTTCTGATTTCCAGCCTTAGACAAATATTATAGACGCTTGAATGATGATCGACTAAT 1500
Qy      481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db      1501 AAAGAAAAATTTGGTAGATGATTTTATTTGGCATTCCTAGCACCAATTACCCATCCAGAGCGA 1560
Qy      501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
Db      1561 CTTGGCAAAACCAAAATTTCAAATTTGAGTATCTGGAAGACGAAGTTCGTATTTGCTCAATTA 1620
Qy      521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAsnGluHisAspIleIleSerAsp 540
Db      1621 GCTGATAAGTATACACAGTCAGATGGTTACATTTTTTGATGAAACATGATATATATCATGTAT 1680

```

```
Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTsrPilleGlyLysAspSer 560
Db 1681 GAAGGAGATGCATATGTAAACCCCTCATATGGCCCATAGTCTAGTGGATGGAAAGATAGC 1740
Qy 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyLysLeuPro 580
Db 1741 CTTTCTGATAGAAAGAGTTGCAGCTCAGCCTACTATAAGAAAAAGGTATCTACCT 1800
Qy 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaAlaLeuTyr 600
Db 1801 CCATCTCCAGACGCAGATGTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTATAC 1860
Qy 601 AsnArgValLysGlyGluLysArgLysProLeuValArgLeuProTyrMetValGluHis 620
Db 1861 AATCGTGTGAAGGGGAAAGAGTTCACCTGCTTCCATATATGTTGAGAT 1920
Qy 621 ThrValGluValLysAsnGlyAsnLeuLeuLeuProHisLysAspHisTyrHisAsnLeu 640
Db 1921 ACAGTTGAGTTAAACCGTAAATTGATTATTCCTCATAGGATCATTAACATAATATT 1980
Qy 641 LysPheAlaTyrPheAspAspHisTyrLysAlaProAsnGlyTyrThrLeuGluAsp 560
Db 1981 AAATTTGCTTGTGTTGATCATCACATACAAAGCTCCAAATGGCTATACCTTGGAAAGT 2040
Qy 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
Db 2041 TTGTTTGCACGATTAAGTACTAGCTAGTAGAACACCCCTGACGACGCTCCACATTTCTAATGAT 2100
Qy 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
Db 2101 GGATGGGGCAATCCCACTGAGCATGTGTTAGGCAAGAAAGACCAAGTGAAGTCCAAAT 2160
Qy 701 LysAsnPheLysAlaAspGluLysProValGluLysProAlaGluThrProAlaGluProGluValPro 720
Db 2161 AAGAACTTCAAGCGCATGAAGCAGCAGTAGAGAAACACCTGCTGAGCAGCAAGTCCCT 2220
Qy 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
Db 2221 CAAGTAGAGACTGAAAGAGTAGAAGCCCACTCAAGAGAGCAGAACTTTGCTTGGAAA 2280
Qy 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
Db 2281 GTAACGGATTTAGTCTGAAAGCCAACTGCAACAGAACTCTAGCTGTTTACGAAATAAT 2340
Qy 761 LeuThrLeuGlnIleMetAspAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780
Db 2341 TTGACTCTTCAAAATTATGATAACAATAGTATATCATGCGAAGCAGAAAAAATTACTTGG 2400
Qy 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796
Db 2401 TTGTTAAAGGAAAGTAACTCTTATCTGTAAGTAAGGAAAAAATAAAC 2448

RESULT 7
AF291695 2541 bp DNA linear BCT 15-MAR-2001
LOCUS Streptococcus pneumoniae pneumococcal histidine triad A protein
DEFINITION (phtA) gene, complete cds.
ACCESSION AF291695
VERSION AF291695.1 GI:13345012
KEYWORDS
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 (bases 1 to 2541)
Wizemann,T.M., Heinrichs,J.H., Adamou,J.E., Erwin,A.L., Kunsch,C.,
Choi,G.H., Barash,S.C., Rosen,C.A., Masure,H.R., Tuomanen,B.,
Gayle,A., Brewah,Y.A., Walsh,W., Barren,P., Lathigra,R., Hanson,M.,
Langermann,S., Johnson,S. and Koenig,S.
Use of a whole genome approach to identify vaccine molecules
affording protection against Streptococcus pneumoniae infection
JOURNAL Infect. Immun. 69 (3), 1593-1598 (2001)
MEDLINE 21116976
```

```
11179332
2 (bases 1 to 2541)
Choi,G.H.
Direct Submission
Submitted (01-AUG-2000) Molecular Biology, Human Genome Sciences,
Inc., 9410 Key West Ave., Rockville, MD 20850, USA
Location/Qualifiers
1..2541
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/strain="W4"
/db_xref="taxon:1313"
91..2541
/gene="phtA"
91..2541
/gene="phtA"
/note="phtA"
/codon_start=1
/transl_table=11
/product="pneumococcal histidine triad A protein"
/protein_id="AAK19155.1"
/db_xref="GI:13345013"
/translation="MKINKKYLVSAAALILSVCSYELGLYQARTVKNRVSIDGK
QATKIENLTPEVSKREGINAEQIVIKIDQGYVTSQGHVHYVNGKVPYDAIIEE
LLMKDPNKLKDEDIVNEVGKIVIKDKIYVYLKDAHADNVTKELINRQKQHS
QHREGGTPNDGAFVALARSQGRYITDDGIFNASDIIEBTGDAYIVPHGDHHTYIPKN
ELSGELAAAEAFLSGRGLNSRTYRQNSDNTSRTNWPSVSPGTTNTNNSNN
TSOASQNSDIDSLKQLYKLPISORHVESGLVDFDPAQITSRTAGVAVPHGDHYHF
IPYSQSLERIEARIIPLYRSNHVPDSPEOPSPOPTPEPSPGAPAPKIDSN
SSLSQLVKRGEGYVPEKGLSRVYFAKDLPESETVKNLESKLSKQESVSHLTJAKE
NVAPDQBFYKAYNLLFEAHKALFNKGRNSDFQALDKLRLDENSTNKEKLVDDL
LAFLAPIHTPERLGNQIEYTEDVETVIAQADKYTSDGYIFDEHDIISDEGDAYV
TPHMGHSWIKGDSLKDEKVAQAQYTKKGLPPSPDADVKANPTGDSAAAIYNKVK
GFKPIPLVRLPYMVEHTVGNLIIIPKDRHYNIKFAWFDHTYKAPNGYTLDFL
ATIKYVHPDERPHSNDGWSNRSHVLGKDKHSDPNKNFKADEPVEETPAEPVP
QVETEKVQLKEAEVLAKVYDSSLSKANATETLAGLRNLTILQIMDNNSIWAEEKL
LALLKGSNPSSVSSEKIN"
ORIGIN
Alignment Scores:
Pred. No.: 8,01e-230 Length: 2541
Score: 4163.00 Matches: 796
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.95% Indels: 0
DB: 1 Gaps: 0
US-09-765-271-56 (1-796) x AF291695 (1-2541)
Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db 151 TCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTAT 210
Qy 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 211 ATAGATGGAAAAACAAGCAGCGCAAAACGGAAATTTGACTCCTCATGAGGTAGCAAG 270
Qy 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db 271 CGTGAAGGATCAATCGCTGAGCAATCGTCATCAAGATAACAGCAAGGCTATGCACT 330
Qy 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db 331 TCACATGGCGCACCATCATATTATTATGTTAAGGTTCTCTTATGACGCTATCATCAGT 390
Qy 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db 391 GAAGAAATTACTCATGAAGAATCAAACTAAGCTAAAGATGAGGATATGTTAATGAG 450
Qy 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db 451 GTCAAGGGTGGATGTTATCAAGGTAGATGGAATAATACTATGTTTACCTTAAGGATGCT 510
```

QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnHisSer 140
 Db 511 GCCCAGCGGATAACGTCGTAACAAAGAGGAATCAATCGACAAACAAAGAGCATAGT 570
 QY 141 GlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
 Db 571 CAACATCGTGAAGGTGGAACTCAAGAAACGATGGTGTGCTTGGCTTGGCAGCTTCGCAA 630
 QY 161 GlyArgTyrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
 Db 631 CGACGCTATATACAGATGATGGTATATCTTTAATGCTTCTGATATCATAGAGATACT 690
 QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
 Db 691 GGTGATGCTTATATCGTTCCTCATCGAGATCATTAACCATTAATTCCTAAGNATGAGTTA 750
 QY 201 SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
 Db 751 TCAGCTAGCGAGTTGGCTGCTGCAGAGCCTTCTTATCTGGTTCGAGGAAATCTGTCAAA 810
 QY 221 SerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
 Db 811 TCAGAACCTATCGCCGACAAATAGCGATTAACACTTCAAGAACAAACTGGGTACCTTCT 870
 QY 241 ValSerAsnProGlyThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
 Db 871 GTAAGCAATCCAGGAATACAAATACTTAACACAGCAACACAGCAACTTAACAGTCAA 930
 QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
 Db 931 GCAAGTCAAAAGTAATGACATTTGATGCTCTGTGAACACAGCTCTACAACTGCCCTTGA 990
 QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
 Db 991 CAACGACATGTAGATCTGATGGCTGTCTTTGATCAGCACAAATCACAAAGTCGAACA 1050
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
 Db 1051 GCTAGAGGTGTGAGTGCCACAGGAGATCATTTACCACTTCACTCCCTTACTCTCAATG 1110
 QY 321 SerGluLeuGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340
 Db 1111 TCTGAATGGAAGAACAAATCGCTGCTGATTTATTCCTCTGCTTATCGTTCAAAACCAATGG 1170
 QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGlnProSerProGly 360
 Db 1171 GTACCACTTCAAGCCAGAACACACCAAGTCCACACCGACTCCGAACTTAGTCCAGGC 1230
 QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
 Db 1231 CCGCAACCTGCACCAATCTTAAATAGACTCAAAATCTTCTTTGTTAGTCAAGTGTGA 1290
 QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
 Db 1291 CGAAAGTTGGGAAGATATGATTTGGAAGAAAGGGCACTCTCTGTTATGCTTTGCG 1350
 QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLysLeuSerLysGlnGluSer 420
 Db 1351 AAAGATTTACCACTTCAAGCTTAAATCTTGAAGAGCAAGTTATCAAAACAGAGAGT 1410
 QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
 Db 1411 GTTTCACACTTAACTGCTTAAAGAAAGAAATGTTGCTCCTCGTGACCAAGAAATTTAT 1470
 QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisIleLysAlaLeuPhe***AsnLysGlyArg 460
 Db 1471 GATAAGCATATATCTGTTAATCGAGGCTCAATAAGCCCTTGTGTGNAATAAGGTCGT 1530
 QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
 Db 1531 AATTCTGATTTCAAGCCTTAGCAATATTAAGAACGCTTGAATGATGATCACTAAT 1590
 QY 481 LysGluLysLeuValAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500

Db 1591 AAAAGAAAATTTGGTAGATGATTTATTTGGCATTTCTTAGCACTAATACCATCCAGAGCGA 1650
 QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
 Db 1651 CTTGGCAACCAAAATCTCAAAATTCAGTATATCTGAAGACGAAGTTCGTATTCCTCAATTA 1710
 QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleSerAsp 540
 Db 1711 GCTGATAAGTATACAACTGTCAGATGGTTACATTTTGTATGAACATGATATATCACTGAT 1770
 QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560
 Db 1771 GAAGAGATGATATATGTAACGCTCATATGGCCCATAGTCTCTGATTTGGAAGATAGC 1830
 QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
 Db 1831 CTTTCTGATAAGGAAAAGTTGCGAGCTCAAGCCTATATCTAAAGAAAAGGATCTCTACT 1890
 QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaAlaIleTyr 600
 Db 1891 CCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTATAC 1950
 QY 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
 Db 1951 AATCGTGTGAAGGGGAAAACGAATTCCTCTGCTGACTTCCATATATATGTTGAGCAT 2010
 QY 621 ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle 640
 Db 2011 ACAGTTCAGGTTAAAAACGGTAAATTTGATTTATCTCTAAGGATCATTAACATAATAT 2070
 QY 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660
 Db 2071 AAATTTCTGTTTGTATGATCAACATACAAAGCTCCAAATGCTATACCTTGAAGAT 2130
 QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
 Db 2131 TTGTTTCGAGCATTAAGTACTACGTAGAACACCCCTCAGCAACGTCACATTTCTAATGAT 2190
 QY 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
 Db 2191 GSATGGGCAATGCCAGTGAAGCATGTGTAGGCAAGAAAGACACAGTGAAGATCCAAAT 2250
 QY 701 LysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720
 Db 2251 AAGAATCTCAAGCGGATGAAGACCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCCT 2310
 QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
 Db 2311 CAAGTAGAGACTGAAAAGTAGAAGCCCAACTCAAGAAAGCAGAAAGTTTGTCTGCGAAA 2370
 QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
 Db 2371 GTAACGATTTCTAGTCTGAAAGCCAATGCACAGAACTCTAGCTGTTTACCAATATAT 2430
 QY 761 LeuThrIleGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780
 Db 2431 TTGACTCTTCAATATTAGTAACAATAGTATCATGCGAAGCAGAAAAATTTACTTGGC 2490
 QY 781 LeuLeuLysGlySerAsnProSerSerValSerLysGlyLysIleAsn 796
 Db 2491 TTGTTAAAAGGAAGTAATCCTTCTATCTGTAAGTAAGGAAAAATAAAC 2538

RESULT 8
 AR218862
 LOCUS
 DEFINITION
 Sequence 94 from patent US 6420135.
 AR218862
 ACCESSION
 AR218862.1 GI:23319796
 VERSION
 AR218862.1
 KEYWORDS
 Unknown.
 SOURCE
 Unknown.
 ORGANISM
 Unclassified.


```

Db 4914 ACAGTTGAGGTTAAACCGGTAATTTGATTATCTCTAATAGGATCATTACCAATAAT 4973
QY 641 LysPheAlaTrpPheAspAspHisThrTyrLysAlaProenGlyTyrThrLeuGluAsp 660
Db 4974 AAATTTCTTGGTTTGAATGATCACACATACAAAGCTCAAATGGCTATACCTTGAAGAT 5033
QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
Db 5034 TTGTTTGGCAGATTAACTACTAGTAAACACCCCTGACCAACCTCCACATTTCTAATGAT 5093
QY 681 GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
Db 5094 GGATGGGCAATGCCAGTGAAGCATGTGTAGGCAAGAAAGACACACAGTGAAGATCCAAT 5153
QY 701 LysAsnPhelLysAlaAspGluGluProValGluThrProAlaGluProGluValPro 720
Db 5154 AAGAATCTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCCT 5213
QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
Db 5214 CAAGTAGAGCTGAAAAGTAGAGGACCACTCAAGAACGACAGATTTTGGCTTGGCAAA 5273
QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
Db 5274 GTAAACGATTCTAGTCTGAAAGCAATGCAACAGAACTCTAGCTGGTTTACGAAATAAT 5333
QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780
Db 5334 TTCACTCTTCAAATTTATGGATAACAATAGTAGTATCATGGCAGAACGAGAAAATTTACTTGC 5393
QY 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleLeu 796
Db 5394 TTGTTAAAGGAAGTAACTCTTCACTCTGTAAGTAAGGAAAAATAAAC 5441

RESULT 9
BD003774
LOCUS 8195 bp DNA linear PAT 31-JAN-2002
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION BD003774
VERSION BD003774.1 GI:18631735
KEYWORDS JP 2001501833-A/94.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 8195)
AUTHORS Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Bara,S.C., Fannon,M. and Dougherty,B.A.
TITLE Polynucleotide of Streptococcus pneumoniae and sequence
JOURNAL Patent: JP 2001501833-A 94 13-FEB-2001;
COMMENT HUMAN GENOME SCIENCES INC
OS Unidentified
PN JP 2001501833-A/94
PD 13-FEB-2001
PF 30-OCT-1997 JP 1998520718
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI STEVEN C BARASH.
PI MICHAEL FANNON, BRIAN A DOUGHERTY
PC C12N15/09, A01K67/027, C07K14/315, C07K16/12, C12N1/15, C12N1/19, C12N1/21,
PC C12N5/10, C12P21/02, C12Q1/68, G06F17/30, C12N15/00, C12N5/00, PC G06F15/40
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..8195 /organism='Unidentified'.
FT Location/Qualifiers
1..8195
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'
FEATURES
Source

```

ORIGIN

```

Alignment Scores:
Pred. No.: 3,48e-229 Length: 8195
Score: 4163.00 Matches: 795
Percent Similarity: 99.87% Conservative: 0
Best Local Similarity: 99.87% Mismatches: 1
Query Match: 99.95% Indels: 0
DB: 6 Gaps: 0

US-09-765-271-56 (1-796) x BD003774 (1-8195)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db 3054 TCTTACAGATTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTGTTCCTAT 3113
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 3114 ATAGATGGAAAAACAGCGACGCAAAAACCGAATTTGACTCTCTGATGAGGTTAGCAAG 3173
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db 3174 CGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCAT 3233
QY 61 SerHisGlyAspHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db 3234 TCACATGGCGACCACTATCATTTATTACATGGTAAGGTTTCTTATGACGCTATCATCAGT 3293
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db 3294 GAAGAATTACTCATGAAGATCCAACTATPAAGCTAAAGATGAGGATATTTGTAATGAG 3353
QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db 3354 GTCAAGGTTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGGATGCT 3413
QY 121 AlaHisAlaAspAsnValArgThrLysGluLysLeuAsnArgGlnLysGlnGluHisSer 140
Db 3414 GCCCAGCGGATTAACCTCGTACAAAAGAGGAATCAATCGACAAAACAAAGACGATAGT 3473
QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyValAlaValAlaLeuAlaArgSerGln 160
Db 3474 CAACATCGTGAAGTGGAACTCCAGAAACGATGGTGTCTGTTCCTTGGCACGTTTCGAA 3533
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db 3534 GGACGCTATACATACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 3593
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db 3594 GGTGATGCTTATATCTGTTCTCATGGAGATCATTAACCATTAACCTTAAGATGAGTTA 3653
QY 201 SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db 3654 TCAGCTAGCGAGTTGGCTGCTCGAGAGGCTTCCTATCTCTGTCGAGAAATCTGTCAAAAT 3713
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
Db 3714 TCAAGAACCTATCGCGCAGCAAAATAGCGATAACACTTCAAGAACCAAACTGGGTACCTTCT 3773
QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db 3774 GTAAGCAATCCAGGAACCTACAATACTAACACAGCAACCAACAGCAACTAACAGTCAA 3833
QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
Db 3834 GCAAGTCAAGTAATGACATTGATGATCTCTTTGAAACAGCTCTACAACCTGCTTTGAGT 3893
QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db 3894 CAACGACATGTAGAATCTGATGGCCTTGTCTTTTGTATCCAGCACAAATCACAAAGTCAACA 3953
QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320

```

Db 3954 GCTAGAGGTGTCAGTGCACACGAGATCATTTACACATTCATCCCTTACTCTCAATG 4013
 Qy 321 SerGluLeuGluArgIleAlaArgIleProLeuArgTyrArgSerAsnHisTyrp 340
 Db 4014 TCTGAATTGGAAGAACGAATCGCTCGTATTATTCCTCCCTTCGTTATCGTTCAACACATTGG 4073
 Qy 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
 Db 4074 GTACACGATTCAAGCCAGAACCAACAGTCCACCGACTCCGGAACTTAGTCCAGGC 4133
 Qy 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
 Db 4134 CCGCAACTGCACCAATCTTAAATAGACTCAATCTCTTTGGTTAGTCAGCTGGTA 4193
 Qy 381 ArgLysValGlyGluGlyTyrValPheGluLysGlyIleSerArgTyrValPheAla 400
 Db 4194 CGAAAGTTGGGAAGGATATGTATTCAGAAAGGGCATCTCTCGTTATGTCTTTGCG 4253
 Qy 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnLys 420
 Db 4254 AAGATTATACCATCTGAACACTGTAAATACTTGAAGCAAGTTATCAAAACAAGAGAGT 4313
 Qy 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
 Db 4314 GTTTCACACACTTAACTGCTAATAAGAAATGTTGCTCTCGTGACCAAGAAATTTAT 4373
 Qy 441 AspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
 Db 4374 GATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCTTGTGTTGAATAAAGGGTCGT 4433
 Qy 461 AsnSerAspPheGlnAlaLeuAspLysLeuGluArgLeuAsnAspGluSerThrAsn 480
 Db 4434 AATTCGTGATTTCCAGCCCTTAGACAAATTTAGAACGCTTGATGATGATGATGATGAT 4493
 Qy 481 LysGluLysLeuValAspAspLeuAlaPheLeuAlaProIleThrHisProGluArg 500
 Db 4494 AAGAAATAATGGTAGATGATTTATGTCATTCCTAGCACCAATTTACCATCCAGAGCGA 4553
 Qy 501 LeuGlyLysProAsnSerGlnIleGlyTyrThrGluAspGluValArgIleAlaGlnLeu 520
 Db 4554 CTTGGCAACCAAAATTTCTCAATTTAGTATATCTAGAACGAAAGTTGCTGCTCAATTA 4613
 Qy 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleSerAsp 540
 Db 4614 GCTGATAAGTATCAACGTCAGATGTTACATTTTGTATGACATGATATATCACTGAT 4673
 Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560
 Db 4674 GAAGGAGATGCATATGTAAGCGCTCATATGGGCCATAGTCACTGGATTGGAAGATAGC 4733
 Qy 561 LeuSerAspLysGluLysValAlaGlnAlaTyrThrLysGluLysGlyLeuPro 580
 Db 4734 CTTTCTGATAAGGAAAAAGTTCCAGCTCAAGCTTATCTAAAGAAAAAGGTATCTTACT 4793
 Qy 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaAlaIleTyr 600
 Db 4794 CCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAGTGCACGACATTTAC 4853
 Qy 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
 Db 4854 AATCGTGTGAAGGGGAAAAAGCAATTCACCTCGTTCCGACTTCCATATATGTTGAGCAT 4913
 Qy 621 ThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIle 640
 Db 4914 ACAGTTGAGTTAAAGACCGTAAATTTGATTATTTCTCATGAAGATCATACCATATAT 4973
 Qy 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660
 Db 4974 AATTTGCTTGGTTGATGATCACATACACATAAGCTCCAAATGGCTATACCTTGGAGAT 5033
 Qy 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680

Db 5034 TTGTTTCGCGAGTAAAGTACTAGTAGAACACCTCGACGACCTCCACATTTCTAATGAT 5093
 Qy 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
 Db 5094 GGATGGGCGCAATGCCAGTGGCATGTCTTAGCGAAGAAAGACACACAGTGAAGATCCAAAT 5153
 Qy 701 LysAsnPhelLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720
 Db 5154 AAGAACTTCAAGCGGATGAAGAGCCAGTGAAGAAACACCTGCTGAGCCAGAAAGTCCCT 5213
 Qy 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
 Db 5214 CAAGTAGAGACTGAAAGAGTAGAAGCCCACTCAAGAAAGCAGAAAGTTTCTTTCGGA 5273
 Qy 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
 Db 5274 GTAACGATTTCTAGTCTGAAGCAATGCAACAGAACTCTAGCTGTTTACGAATAAT 5333
 Qy 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780
 Db 5334 TTGACTCTTCAAAATATGATGATAACATAGTATCATGCGAAGCAGAAAAATTTACTTGG 5393
 Qy 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796
 Db 5394 TTGTTAAAGAGAGTAATCTTCTCATCTGTAAGTAAGAAAAAATAAAC 5441

RESULT 10
 LOCUS AE007418/1 10256 bp DNA linear BCT 31-AUG-2001
 DEFINITION Streptococcus pneumoniae TIGR4 section 101 of 194 of the complete genome.
 ACCESSION AE007418 AE005672
 VERSION AE007418.1 GI:14972649
 KEYWORDS
 SOURCE Streptococcus pneumoniae TIGR4
 ORGANISM Streptococcus pneumoniae TIGR4
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 REFERENCE 1 (bases 1 to 10256)
 Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T., Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M.
 Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*
 Science 293 (5529), 498-506 (2001)
 JOURNAL
 MEDLINE 21357209
 PUBMED 11463916
 REFERENCE 2 (bases 1 to 10256)
 Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Umayam, L.A., White, O., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Gesuwan, P., Hickey, E.K., Holt, I.E., Loftus, B.J., Ujwal, M.L., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M.
 Direct Submission
 Submitted (29-JUN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
 JOURNAL
 FEATURES
 Location/Qualifiers
 1..10256
 /organism="Streptococcus pneumoniae TIGR4"
 /mol_type="genomic DNA"
 /strain="TIGR4"
 /db_xref="taxon:170187"
 complement(100..702)
 /gene="SP1170"

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVallysgluAsnAsnArgValSerTyr 20
 Db 7736 TCTTACGAGTGGGACTGTATCAAGCTPAGAACGGTTAAGGAAATAATATCGTTTTCCTAT 7677

QY 21 IleAspGlyGlyGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
 Db 7676 ATAGATGGAAACAAAGCGACGCAAAACCGAGAAATTTGACTCTCTGATGAGGTAGCAAG 7617

QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
 Db 7616 CGTGAAGGAATCAATCGTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCACT 7557

QY 61 SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
 Db 7556 TCACATGGCGACCACTATCATATTATCAATGGTAAGGTTCTCTATGACGCTATCATCAGT 7497

QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
 Db 7496 GAAGAATTACTCAAGAAGATCCAACTATAAGCTAAAGATGAGGATATTGTTAATGAG 7437

QY 101 VallysglyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
 Db 7436 GTCAAGGCTGTATGTTATCAAGGTAGATGGAATAATATATGTTTAAAGGATGCT 7377

QY 121 AlaHisAlaAspAsnValArgThrLysGluLileAsnArgGlnLysGlnLysSer 140
 Db 7376 GCCCAGCGGATAAGCTCGTACAAAGAGGAATCAATCGACAAAAACAAGAGCATAGT 7317

QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
 Db 7316 CAACATCGTGAAGTGGAACTCCAGAAACGATGTGTGTTGCTTGGCAGTTCGCAA 7257

QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
 Db 7256 GGAGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGTATATCATAGAGATACT 7197

QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlnLeu 200
 Db 7196 GGTGATGCTTATATCGTTCTCATGGAGATCAATACCATTCATCTTCAAGNAGTGA 7137

QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
 Db 7136 TCAGCTAGCGAGTGGCTGCTGCAGAGCTTCTTATCTGTGTCGAGGAATCTGTCAAA 7077

QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
 Db 7076 TCAAGAACCTATCGCCGACAAATAAGCGATAACACTTCAAGAACAACTGGGTACTTCT 7017

QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
 Db 7016 GTAAGCAATCCAGGAACATAAATACTAACACAGCAACACAGCAACACTAACAGTCA 6957

QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
 Db 6956 GCAAGTCAAAAGTAATGACATTGATAGTCTCTTGAACACAGCTCTACAACTGCTTGTAGT 6897

QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
 Db 6896 CAACGACATGTAGATCTGATGGCTTGTCTTGTATCCAGCAACAAATCAAGTGCACAA 6837

QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
 Db 6836 GCTAGAGGTGTCAGTGCACACGAGATCATACCACTCATCTCTACCTCAATG 6777

QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340
 Db 6776 TCTGAATCGAAGACGAATCGCTGATATTATTTCCCTTCGTTATCGTTCAAAACCATG 6717

QY 341 ValProAspSerArgProGluGlnProSerProGlnProGluProSerProGly 360
 Db 6716 GTACCAAGATTCAAGGCCAGAACCAACCAAGTCCAAACCGACTCCGAACCTAGTCCAGGC 6657

QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
 Db 6656 CCGCAACCTGCACCAAAATCTTAAATAGACTCAAAATCTTCTTTGTTAGTACAGTGGTA 6597

QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
 Db 6596 CGAAAGTTGGGAAGGATATGATTTCGAAGAAAGGGCATCTCTCGTTATGTTCTTTGCG 6537

QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLysLeuSerLysGlnLysSer 420
 Db 6536 AAGATTACATCTGAACCTGTTAAATCTTGAAGCAAGTTATCAAAACAAGAGAGT 6477

QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
 Db 6476 GTTTCACACACTTTAACTGCTAAAAAAGAAATTTGCTCTCTCTGACCAAGAAATTTAT 6417

QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
 Db 6416 GATAAAGCATATATCTGTTAACTGAGGCTCATTAAGCCTTGTGTAATAAGGTCGT 6357

QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
 Db 6356 AATTCTGATTTCCAAAGCTTAGACAAATTAATGACACGCTTGAATGATGAATCAGCTAAT 6297

QY 481 LysGluLysLeuValAspAspLeuAlaPheLeuAlaProIleThrHisProGluArg 500
 Db 6296 AAAGAAAAATGCTAGATGATTTATGTCATCTCTAGCACCAATTAACCATCCAGAGCGA 6237

QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
 Db 6236 CTTGGCAAAACCAAAATCTCAAAATGAGTATATCTGAAGCAAGATTCGTATTGCTCAATTA 6177

QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
 Db 6176 GCTGATAAGTATACACGTCAGATGGTTTACATTTTGTATGAACATGATATATTAATCAGTGAT 6117

QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560
 Db 6116 GAAGGAGATGCATATGTAACGCTCATATGGGCCATAGTCTCATCTGGATTGGAAAAAGATAGC 6057

QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
 Db 6056 CTTTCTGATAGGAAAAAGTTGCGACTCAAGCTATATCTAAAGAAAAAGGTATCTTACT 5997

QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaAlaIleTyr 600
 Db 5996 CCATCTCCAGCGCAGATGTTAAAGCAAAATCCAACTCGAGATAGTGCAGCAGCTATTTAC 5937

QY 601 AsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
 Db 5936 AATCGTGTGAAAGGGGAAAAACGAATCCACTCGTTCGACTTCCATATATGTTGAGCAT 5877

QY 621 ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle 640
 Db 5876 ACGTTTGAGTTAAACCGTAATTTGATTATCTCTCATAGGATCAITTAACATAATTT 5817

QY 641 LysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660
 Db 5816 AATTTGCTTGTGTTGATGATCACACATACAAAGCTCCAAATGGCTATATACCTTGAAGAT 5757

QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
 Db 5756 TTGTTTGGCAGCAATTAAGTACTACGTAGAACACCTTGAGAACGTCACATCTTAATGAT 5697

QY 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
 Db 5696 GAATGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACACACAGTCAAGATCCAAAT 5637

QY 701 LysAsnPheLysAlaAspGluProValGluThrProAlaGluProGluValPro 720
 Db 5636 AAGAACTTCAAAAGCGGATGAAGAGCCAGTAGAGAAACACCTGCTGAGCGAGAGTCCCT 5577

QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740


```

Db      5576  CAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAGAGAGCAGAGATTTCCTGCGGAA 5517
QY      741  ValThrAspSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
Db      5516  GTAACGGATTCCTAGCTGAAAGCCAAATGCAACAGAAACTCTAGCTGGTTTACGAAATAAT 5457
QY      761  LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780
Db      5456  TTGACTCTTCAAAATATGATAACAATAGTATCATGGCAGAGCAGAGAAATTTACTTGGC 5397
QY      781  LeuLeuLysGlySerAsnProSerValSerLysGluLysIleAsn 796
Db      5396  TTGTTAAAGGAAGTAATCCTCTCATCTGTAAGTAAGCAAGAAAAATAAAC 5349

RESULT 11
AX571763/c
LOCUS   AX571763 349980 bp DNA linear PAT 29-NOV-2002
DEFINITION Sequence 4982 from Patent WO02077021.
ACCESSION AX571763
VERSION   AX571763.1 GI:26003955
KEYWORDS .
SOURCE   Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
          Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
          Streptococcus.
REFERENCE
1. Masignani, V., Tettelin, H. and Fraser, C.
   TITLE   Streptococcus pneumoniae proteins and nucleic acids
   JOURNAL Patent: WO 02077021-A 4982 03-OCT-2002;
   Chiron Spa (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
   Location/Qualifiers
     1. 349980
        /organism="Streptococcus pneumoniae"
        /mol_type="unassigned DNA"
        /db_xref="taxon:1313"
        /note="seq 4979 too long: 2,162,598 bases-replaced by
        following seq:-seq 4979: from 0.000.001 to 0.349.980-seq
        4980: from 0.300.001 to 0.649.980-seq 4981: from 0.600.001
        to 0.949.980-seq 4982: from 0.900.001 to 1.249.980-seq
        4983: from 1.200.001 to 1.549.980-seq 4984: from 1.500.001
        to 1.849.980-seq 4985: from 1.800.001 to 2.149.980-seq
        4986: from 2.100.001 to 2.162.598"

ORIGIN
Alignment Scores:
Pred. No.: 3 84e-227 Length: 349980
Score: 4163.00 Matches: 795
Percent Similarity: 99.87% Conservative: 0
Best Local Similarity: 99.87% Mismatches: 1
Query Match: 99.95% Indels: 0
DB: 6 Gaps: 0

US-09-765-271-56 (1-796) x AX571763 (1-349980)

QY      1  SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnArgValSerTyr 20
Db      106949  TCTTACGAGTTGGGACTGATCAAGCTAGAACGGTTAAGGAAAAATATCGTTTCCTAT 106890
QY      21  IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db      106889  ATAGATGGAACAAAGCGACGCAACAAAAACGGAGATTTTGAATCTCTGATGAGGTAGCAAG 106830
QY      41  ArgGluGlyLysAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db      106829  CGTGAGGAAATCAATGCTGAGCAATCGTCATCAAGATTAACAGCAACGAGCTATGTCAT 106770
QY      61  SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db      106769  TCACATGGCGACCACTATCATTTATACATGTTAGGTTTCCTTATGACGCTATCATCAGT 106710
QY      81  GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100

```

```

Db      106709  GAAGAAATTACTCATGAAGATCCAAACTATAAGCTAAAGATGAGGATATTGTTAATGAG 106650
QY      101  ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db      106649  GTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAACTATGTTTACCTTAAAGATGCT 106590
QY      121  AlaHisAlaAspAsnValArgThrLysGluLysLeuAsnArgGlnLysGlnGluHisSer 140
Db      106589  GCCCACCGGATACACGTCCTGTACAAAAGAGAGAAATCAATGCACAAAACCAAGAGCATAGT 106530
QY      141  GlnHisArgGluGlyGlyThrProArgAsnAspGlyValAlaLeuAlaLeuAlaArgSerGln 160
Db      106529  CAACATCGTGAAGTGGAACTCCAGAAACGATGGTGTGTTGCTTGGCACGTTTCGCA 106470
QY      161  GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db      106469  GGAACGCTTATACAGATGATGGTTATATCTTTAATGCTTCTCATATCATAGAGGATACT 106410
QY      181  GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db      106409  GGTGATGCTTATATCGTTCCTCATGGAGATCATTCACATTACATTCCTAAGAATGAGTTA 106350
QY      201  SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db      106349  TCAGCTAGCGAGTTGGTCTGCAGAGCCCTTCTATCTGGTCGAGAAATCTGTCAAAT 106290
QY      221  SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
Db      106289  TCAAGAACCTATCGCGCAAAAATAGGATAACACTTCAAGAACAACTGGGTACCTTCT 106230
QY      241  ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db      106229  GTAAGCAATCCAGGAATACTAAATACTAACACAAAGCAACAACTAACAGTCA 106170
QY      261  AlaSerGlnSerAsnAspIleAspSerLeuLysGlnLeuTyrLysLeuProLeuSer 280
Db      106169  GCAAGTCAAGTAATGACATGTAGTCTCTTGAACACGCTCTACAAACTGCTTTGAGT 106110
QY      281  GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db      106109  CAACGACATGTAGATCTCATGTCCTTGTCTTGTATCCAGCACAAATCAACAGTCCAA 106050
QY      301  AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db      106049  GCTAGAGGTGTCAGTGCACGAGATCATTAACACTTCTATCCCTTACTCTCAATG 105990
QY      321  SerGluLeuGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340
Db      105989  TCTGAATTCGAAGACGAATCGCTCGTATATTCCTCTTTCCTTATCGTTTCAACACCATGG 105930
QY      341  ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db      105929  GTACCAGATTCAAGGCCAGAACCAACAGTCCACACCGACTCCGGAACCTAGTCAGGC 105870
QY      361  ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
Db      105869  CGCAACCTCGCACCAAAATCTTAAATAAGATAGACTCAAAATCTCTTGTAGTCAGTGGTA 105810
QY      381  ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
Db      105809  CGAAAAGTTGGCAAGGATATGTAATTCGAAGAAAAGGGCATCTCTCGTTATGCTTTGGC 105750
QY      401  LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
Db      105749  AAAGATTACCATCTGAAACTGTTAAAAATCTTGAAGCAAGTATATCAAAACAGAGAGT 105690
QY      421  ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
Db      105689  GTTTCACACACTTAACTGCTAAAAAGAAAATGTTGCTCTCGTGCACCAAGAAATTTAT 105630
QY      441  AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db      105629  GATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGAAAAATAAGGGTCGT 105570

```



```

/gene="spr1058"
/codon_start=1
/transl_table=11
/label=spr1058
/product="Hypothetical protein"
/protein_id="AAK99862.1"
/db_xref="GI:15458680"
/translation="MNLKRVTDPSDKSIINRPVIVEGAYNAALQAGAEISVIQTQ
LAELKINK"
complement(2151..2306)
/gene="spr1059"
/codon_start=1
/transl_table=11
/label=spr1059
/product="Hypothetical protein"
/protein_id="AAK99863.1"
/db_xref="GI:15458681"
/translation="MMESIGIGLVSHSHKIAEGVVELISKVADVPITVVGTEGG
GIGTSFD"
complement(2542..2613)
/rpt_family="RUP element"
/rpt_type="dispersed"
/rpt_unit="rupA"
complement(2725..5292)
/gene="phpA"
/notes="synonym: spr1060"
complement(2725..5292)
/gene="phpA"
/codon_start=1
/transl_table=11
/label=spr1060
/product="Histidine Motif-Containing protein"
/protein_id="AAK99864.1"
/db_xref="GI:15458682"
/translation="MNOIYLKREMRKINKYLAGSVATLVLSVCAYELGLHQATVK
ENRVSVIDGKQATQKENTLPDEVSREKGINAEQIVIKITDOGYVTSBGDHYHYNG
KVPDAIISSELLMKDPNYQLKDEDIISIKGGVIVKDGKYVYLKDAHADNVRTK
EINRQKHSHQREGTFRNDGAVALSQGYTTDDGYIFNADSIIEDTDGAYIVP
HGDHYIYPKNEISAGELAAAFSLSGRGLNSRTYRQNSDNTSRNTWVPSVNP
TTVNTNNSNTSQAQSDNDISLLQLYKPLSRHVESDGLIFDPAQITRTANG
VAPNPGHFIPIYQSLPLEKLARIIPLYRSNHWVPSRPEQSPQSPPTSPSP
KQESLHKLGAKTDLPSDRDEFYNKAYDILLARIHODLONKGRQVDFEALNDLER
LDVSSDKVLVDIILAFIPIRPERLGPNAQITVTDDEIQVAKLAGYTTEDGYI
FPRDITSDGDAYVTPMTHSHWIKDSLSERAAQAAYAKEGLTPP2DHDQSG
NTEAKGAEALYNRVAAKVPDRMPINLQITVEVKNGLIIPHYDHYNIKEPWFDE
GLYEAQKGSLELLARVYVEHPNRPDSNGFNASDHYVQRNKGQADNTQTEKP
NEEKQTEKPEBETPREEPQSEKPSKPTEPEEPEEPEEPEEPEEPEEPEEPEE
KLREABDLGKIQNPFIKSNKAKETLTGLKNLLFGTQDNNTIMAEAEKLLALLKESK"
complement(5417..7903)
/gene="phtA"
/notes="synonym: spr1061"
complement(5417..7903)
/gene="phtA"
/codon_start=1
/transl_table=11
/label=spr1061
/product="Pneumococcal histidine triad protein A
precursor"
/protein_id="AAK99865.1"
/db_xref="GI:15458683"
/translation="MQLETSNRKRVSMKINKYLVGSAAALLVSCVSELGLYQARTV
KENRVSYIDGKQATQKENTLPDEVSREKGINAEQIVIKITDOGYVTSBGDHYHYN
KVPYDAIFSELLMKDPNYKLKDEDIINEVKGYYIVKDGKYVYLKDAHADNVRTK
KEEINRQKHSHQREGTFRNDGAVALSQGYTTDDGYIFNADSIIEDTDGAYIVP
HGDHYIYPKNEISAGELAAAFSLSGRGLNSRTYRQNSDNTSRNTWVPSVNP
GTTNTNNSNTSQAQSDNDISLLQLYKPLSRHVESDGLIFDPAQITRTAR
GVAPPHGDHYHPIYQSOMSELEERARIIPLRYSNHWVPSRPEQSPQSPPTSPSP
POPAPNLKIDSNSILVSQIVKRGYVFEKISRYVFAKDLPSFTVKNLESLSKQ
ESVSHTLTKENVAPROEYFDKAYNLLTAHKALFENKGRNSDFQALDKLERLND
ESTNKEKLVDDLLAFIPIRPERLGPNSQIETVEDEVRIQAADKYVTSBGYIFDE

```

```

gene
HDIISDEGDAYVTPMHGSHWIGKDSLCKEKVAQAQYTKKGILPPSPDADKANPT
GDSAAAIYNRVYKGERIIPLVRLPYMVEHTVEVKNGLIIPKHDXHNIKFAFDHHTY
KAPNGYTLIEDLFATIKYVVEHDPDRPHSNDGWNASEHVLGKDHSEDPKNKPKADEE
PVEETPARPEVPEQVETEKVQALKEAEVLAKVTDSLSKANATETIAGLNNLITQIM
DNNSIMAEAEKLLALLKSNFSSSVSEKIN"
complement(8122..9855)
/gene="ptsI"
/notes="synonym: spr1062"
complement(8122..9855)
/gene="ptsI"
/EC_number="2.7.3.9"
/codon_start=1
/transl_table=11
/label=spr1062
/product="PTS enzyme I"
/protein_id="AAK99866.1"
/db_xref="GI:15458684"
/translation="MTEMLGIAASDGVAVAKAYLLVQPDLSFETITVEDTNABEARL
DAIQAODELSVIREKAVGTGEEAAQVFDALHMLVADPEMISOIKETIRAKKVNAE
AGLEKVTDMFTITFEQMEDNPMQBERAADIRDTKRVLANLLGKLPNPASINEEVIJ
IAHDLTPSDTAQLDKNFVKAFVTNIGGRTSHSAIWAETLEIAAVLTGNNITEIKDGD
ILAVNGITGEVIINPTDQAEAFKAAGEAYAKQAEKALLKDAQTVTADKHFEALAN
IGTPKQVEGVNNGAEAVGLYRTEFLMDSQDPTFEDQYEAQVLEGMNGKFPVVR
TMDIGGKELPYFDMPEHMPFLGFRALRISISSETGDAMFTQIRALIRASVHQLRI
MFPWVALLKEFRAAKAVFDEKANKLAEVAVADNIQVGINIEIPAAAMLADQFAKEV
DFFSIGTNDLIQYTMADRMNEQVSYLYQVYNPSILRLINNVIKAAHAEGKAWGCGE
MAGDQAVPLLVGMGLDFSMATSVLTRSLMKKLDOTAKMEEYANRALTECSIMEEV
LELQKEYVNF"
complement(9861..10124)
/gene="ptsH"
/notes="synonym: spr1063"
complement(9861..10124)
/gene="ptsH"
/codon_start=1
/transl_table=11
/label=spr1063
/product="Histidine-containing phosphocarrier protein of
the PTS"
/protein_id="AAK99867.1"
/db_xref="GI:15458685"
/translation="MASHQFHVVAETGIIHARPATLLVQVTSKFPASDITLEYKGSVNL
KSMGVMSLVGGQADVITSAEGADADDAIAAISETMEKEGLA"

```

ORIGIN

Alignment Scores:	7,87e-229	Length:	10320
Pred. No.:	4159.00	Matches:	794
Score:	99.75%	Conservative:	0
Percent Similarity:	99.75%	Mismatches:	2
Best Local Similarity:	99.86%	Indels:	0
Query Match:	1	Gaps:	0
DB:	US-09-765-271-56 (1-796) x AE008479 (1-10320)		

QY	1	SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnArgValSerTyr	20
Db	7807	TCTTACGAGTTGGAGCTGTATCAAGCTAGACGGTTAAGGAAAAATATCGTGTTCCTAT	7748
QY	21	IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys	40
Db	7747	ATAGATGGAAAAACAAGCGACGCAAAAAACGGAGAAATTTGACTCTCTGAGGTAGCAAG	7688
QY	41	ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr	60
Db	7687	CGTGAAGGAATCAATGCTGAGCAATTCGTATCAAGATAACAGACCAAGGCTATGTCT	7628
QY	61	SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer	80
Db	7627	TCACATGCGCACCATCATCATTTTACAAATGGTAAAGTTTCCTTATGATCCGATTTTCA	7568
QY	81	GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu	100
Db	7567	GAAGAAATCTCATGAAAGATCCAACTATAAGCTAAAAGATGAGGATTTGTTTAATGAG	7508

QY	101	ValIysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla	120	AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn	480
QY	121	AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer	140	LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg	500
QY	141	GlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln	160	LeuGlyLysProAsnSerGlnIleGlyThrGluAspGluValArgIleAlaGlnLeu	520
QY	180	GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleLeuGluAspThr	200	GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisThrIleGlyLysAspSer	560
QY	220	SerAlaSerGluLeuLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn	240	ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr	600
QY	260	GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu	280	ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle	640
QY	300	GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr	320	GlyThrProGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn	700
QY	340	SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr	360	ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly	720
QY	380	ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal	400	GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys	740
QY	420	ArgLysValGlyGluGlyTyrValPheGluGluLysGlyLysSerArgTyrValPheAla	440	ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn	760
QY	460	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	480	GTAACGGGATTCCTAGTCTGAAAGCCCAATGCCACAGAACTCTAGCTGGTTACGAAATAAT	528
QY	500	LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer	520	LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla	780
QY	540	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	560	LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn	796
QY	580	GTAACGGGATTCCTAGTCTGAAAGCCCAATGCCACAGAACTCTAGCTGGTTACGAAATAAT	600	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	620	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	640	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	660	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	680	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	700	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	720	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	740	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	760	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	780	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	800	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	820	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	840	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	860	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	880	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	900	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	920	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	940	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	960	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	980	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	1000	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	1020	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	1040	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	1060	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	1080	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	1100	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	1120	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	1140	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	1160	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	1180	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	1200	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	1220	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	1240	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	1260	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	1280	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	1300	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	1320	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	1340	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	1360	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	1380	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	1400	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	1420	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	1440	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468
QY	1460	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	1480	TTGACTCTTCAAAATTTATGATAACAATAGTATATATGCGAGAGCAGAGAAATTAATCTTGGC	5468

```

AL449937.1 GI:11545162
VERSION HTG, HTGS PHASE2.
KEYWORDS Streptococcus pneumoniae
SOURCE Streptococcus pneumoniae
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE 1
AUTHORS Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A.,
Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
Francesco, M., Polissi, A., Buell, G., Feger, G., Garcia, E., Peitsch, M.
and Garcia-Bustos, J.F.
TITLE Annotated draft genomic sequence from a Streptococcus pneumoniae
type 19F clinical isolate
JOURNAL Microb. Drug Resist. 7 (2), 99-125 (2001)
MEDLINE 21335329
PUBMED 11442348
REFERENCE 2 (bases 1 to 20035)
AUTHORS Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A.,
Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
Francesco, M., Buell, G., Feger, G., Garcia, E., Peitsch, M. and
Garcia-Bustos, J.F.
TITLE Direct Submission.
JOURNAL Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
Severo Ochoa 2, 28760 Tres Cantos, SPAIN
COMMENT * NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
location/Qualifiers
1..20035
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/serotype="19F"
/db_xref="taxon:1313"
/clone="G54"
ORIGIN
Alignment Scores:
Pred. No.: 8,97e-226 Length: 20035
Score: 4112.00 Matches: 787
Percent Similarity: 98.75% Conservative: 0
Best Local Similarity: 98.75% Mismatches: 9
Query Match: 98.73% Indels: 1
DB: 2 Gaps: 0
US-09-765-271-56 (1-796) x SPNEU1915 (1-20035)
QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db 17071 TCTTACGAGTTGGAGCTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTAT 17130
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 17131 ATAGATGGAAACACACGACGCAAAACGGAGAAATTTGACTCTCGATGAGGTAGCAAG 17190
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db 17191 CQTGAAGGAATCAATGCTGAGCAATCGTCATCAAGATTAACAGCCAGGCTATGTCAT 17250
QY 61 SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db 17251 TCACATGGCGACCATCATATTATCAATGTTAAGGTTCCATATGATCGATTTTCAGT 17310
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db 17311 GAAGAATTAATCATGTAAGATCAAACTATAAGCTAAAGATGAGGATATGTTAATGAG 17370
QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db 17371 GTCAAGGGTGGATATGTTATCAAGGTAGATGTAATAATACTATGTTTACCTTAAGGATGT 17430
QY 121 AlaHisAlaAsnValArgThrLysGluGluIleAsnArgGlnLysGlnHisSer 140
Db 17431 GCCACGCGGATACGTCCTGACAAAAGAGAAATCAATCGACAAAACAAACATAGT 17490
QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
Db 17491 CAACATCGTGAAAGTGGAGCTCCCAAGAAACGATGGTGGGTAGCTTTGACAGCTTCAG 17550
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db 17551 GGAGCTACACCAAGATGATGGTTATATCTTTAATGCTTCTGATATCAATGAGGACAGC 17610
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db 17611 GGTGATGCTTATATCGTCTCTCATGGCAGCAATTAACATTAATCTTAAGAATGAGTTA 17670
QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db 17671 TCAGTAGCGAGTTTGCTGCTGCAGAGGCTTCTCTATCTGGTGGGGAATCTGTCAAT 17730
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
Db 17731 TCAAGAACCTATCGCGACAAATAGCGATAACACTTCAAGAACAACTGGTACCTTCT 17790
QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db 17791 GTAAGCAATCCAGAACTACAAATATCTACACAGCAACCAACAGCAACCTAACAGTCAA 17850
QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
Db 17851 RCAAGTCAAAAGTAATGACATTTGATGCTCTTCAACACAGCTCTACAACTGCTTTGAGT 17910
QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProIleGlnIleThrSerArgThr 300
Db 17911 CAACGMCATGTAGATCTTGATGGCTTTGTTTCGACCCAGCGCAATCAACAGTCAACA 17970
QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db 17971 GCTAGAGGTGTTGCGAGTGCACACGAGATCATACACTTCTATCCCTTACTCTCAATG 18030
QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340
Db 18031 TCTGAATTTGGAGACGAATCGCTCGTATTAATTTCCCTTCGTTATGTTCAACACATGG 18090
QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db 18091 GTACCAGATTCAGGCCAGACCAACCAAGTCCACACCACTCCGGAACCTAGTCCAGGC 18150
QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
Db 18151 CCGCAACCTGCACCAAACTTTAAATAGACTCAATTTCTTTGTTAGTCAGCTGGTA 18210
QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
Db 18211 CGAAAAAGTTGGGGAAGATATGTTATCGAAGAAAGGGGCATCTCTGTTATGTTTGG 18270
QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
Db 18271 AAAGATTTCACCATCTGAACTGTAAATAATCTTGAAGCAAGTTATCAAAACAGAGACT 18330
QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
Db 18331 GTTTCACACACTTAACTGCTTAAAGAAAGAAATGTTGCTCTCGTGACCAAGAAATTTT 18390
QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db 18391 GATAAAGCATATAATCTGTTAACTAGGCTCATAAASCTTGTGTTGAAAATAAGGTCGT 18450
QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThr-As 480
Db 18451 AATCTGATTTCCAAASCTTGGACAAATTTTTTAGAACGCTTGAATGATGATCGACTTAA 18510
QY 480 nLysGluLysLeuValAspLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db 18511 TAAAGAAAAATTGGTAGATGATTTTATGGCATTTCTAGACCAATATCCCATCCAGAGC 18570

```

AUTHORS

Zhang, Y., Masi, A.W., Barniak, V., Mountzouros, K., Hostetter, M.K. and Green, B.A.

TITLE

Recombinant PnpA protein, a unique histidine motif-containing protein from *Streptococcus pneumoniae*, protects mice against intranasal pneumococcal challenge

JOURNAL

Infect. Immun. 69 (6), 3827-3836 (2001)

MEDLINE

21246685

PUBMED

11349048

REFERENCE

2 (bases 1 to 2535)

AUTHORS

Zhang, Y., Masi, A., Barniak, V., Mountzouros, K., Hostetter, M. and Green, B.

TITLE

Direct Submission

JOURNAL

Submitted (25-JAN-2001) Department of Bacteriology, Wyeth Lederle Vaccines, 211 Bailey Road, West Henrietta, NY 14586, USA

FEATURES

Location/Qualifiers

1..2535

Source

/organism="Streptococcus pneumoniae"

/mol_type="genomic DNA"

/db_xref="taxon:1313"

1..2535

gene

1..2535

CDS

1..2535

/gene="pnpA"

/note="histidine motif-containing protein"

/codon_start=1

/transl_table=11

/product="PnpA"

/protein_id="AAK26629.1"

/db_xref="GI:13447094"

/translation="MKINKYLVGSAALILSVCSYELGLYQARTVKNRVSVIDGK

CAQTKENLPDEVSKREGINAEQIVIKIDQGVVTHGHDHYHYNGKVPYDALISEE

LLMKDPNYQLKDEIISEIKGVYIVKDGYYVYLKDAHADHYNRTKEEINRQOEHS

QHRGGPRNDGAVARLQSGRYTDDGYIFNASDIIEDTGDAYIVPHGDHYHYPKN

ELASAEIAAKAFPLSGRGNLSNRTSRNSDNTSRNVPSVSGNPGVTNTNTNNSN

TNSQASQNSDILSKOLYKPLPSORHVESDGLIFDPAQITRTANGVAVPHGDHYH

IPYSLQSLPEKELRIIPLRYRSHNVWDSRPEQSPQSPQSPQSPQSPQSPQSPQSP

NPIDKILKSAVRKVGVDGYVFEENGVEYIIPAKDLSAETAAGIDSKLAKQBSLSHKL

AKTDDLPSSREFRYKAYDLARLHQDLNKKQVDPEALDNLEKLVKQVSSDKVKL

VDDILAFIAPIRHPEKPNQAITTDDDEIQVAKLAKYTTEDGYIFDPRDIYSDEG

DAYVTPMTHSHVTKDLSSEAAKAAVAKKELGTPPSDHDQDSNGTEAKGAEAY

NRVAKKAVPLDRMPYNIQYTVKNGSLIIPHYDVHNIKEFDFEGLYAPKGYSL

EDLLATVKYVYVHEHNERPHSDNGFNGASDHVORNGKQADNTQEKNEKPKQTEKE

EEETPREKPOSEKPEPTEPEEPSEPESEPESEPESEPESEPESEPESEPESEPESE

IQNPPIKSNKAKETITGLKNLLFGTQDNNTIMEAEKLLALLKESK"

ORIGIN

Alignment Scores: 1.17e-176 Length: 2535
Pred. No.: 3236.00 Matches: 620
Score: 83.15% Conservative: 66
Percent Similarity: 75.15% Mismatches: 97
Best Local Similarity: 77.70% Indels: 42
Query Match: 1 Gaps: 4
Db:

US-09-765-271-56 (1-796) x AF340221 (1-2535)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVallysGluAsnAsnArgValSerTyr 20
Db 61 TCTTACGAGTGGGACTGATCACTAGACGCTAAGGAAATAATCGTGTCTTCTAT 120
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 121 ATAGATGAAACCAAGCAGCGCAAAACCGGAGAAATTTGACTCCTCGATGAGGTAGCAAG 180
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db 181 CGTGAAGGATCAATCCCGAACAATCGTCATCAAAATAACAGACCAAGGCTATGTCAT 240
QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleSer 80
Db 241 TCACATGGGACCACTATCATTTATACATGTTAAAGGTTCTTATGACGCTATCATCACT 300
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLysLysAspGluAspGluValAsnGlu 100

QY 500 gLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLe 520
Db 18571 ACTTGGCAAGCCAAATCTCAAATTTGATGATCTACTGAAGACGAAGTTCGTATTGCTCAAT 18630
QY 520 uAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleLeuSerAs 540
Db 18631 AGCTGATGAAGTATACACGTCAGATGTTTCAATTTTGTGATGACATGATATATCATGGA 18690
QY 540 pGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSe 560
Db 18691 TGAAGGAGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 18750
QY 560 rLeuSerAspLysGlnLysValAlaAlaGlnAlaTyrThrLysGlnLysGlyIleLeuPr 580
Db 18751 CCTTCTGATAAGGAAAGATGTCAGCTCAAGCCCTACTAAAGAAAAGATATCCCTACC 18810
QY 580 oProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTy 600
Db 18811 TCCATCTCCAGACGACGATGTTAAAGCAATCCACTGGATAGTGCAGCAGCTATTTA 18870
QY 600 rAsnArgValLysGlyGlnLysArgIleProLeuValArgLeuProTyrMetValGluHi 620
Db 18871 CAATCGTGTAAAGGGAACCAATTCCTCGTTCGACTTCCATATATATGTTGAGCA 18930
QY 620 sThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIl 640
Db 18931 TACAGTTGAGGTTAAACCGGTAAATTTGATTTATTCCTCATAGGATCATTAACCTAACAT 18990
QY 640 eLysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAs 660
Db 18991 TAAATTTGCTTGGTTTGTATGATCACACATACAGCTCCAAATGGCTATATCTTGGAGA 19050
QY 660 pLeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAs 680
Db 19051 TTTGTTTCGACGATTAAGTACTAGTAGAACACCCCTGACGAGCAAGTCCACATTTCTAATGA 19110
QY 680 pGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAs 700
Db 19111 TGGATGGGGCAATGCGCAGTGAGCATGTGTAGGCAAGAACACACACAGTGAAGATCCAAA 19170
QY 700 nLysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPr 720
Db 19171 TAAGNACTTCAAGCGGATGAAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCC 19230
QY 720 oGlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuAlaIly 740
Db 19231 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGCAGAGATTTTGTCTTGGCAA 19290
QY 740 sValThrAspSerSerLysLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAs 760
Db 19291 AGTAACGGATCTAGTCTGAAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAAATAA 19350
QY 760 nLeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAl 780
Db 19351 TTTGACTCTTCAATTTATGATAACAATAGTATCATGCGAGAGCAGAGAAATTTACTTGC 19410
QY 780 aLeuLeuLysGlySerAsnProSerSerValSerLysGlnLysIleAsn 796
Db 19411 GTTGTAAAGGAAGTAACTCTTCTCTCTGTAAGTAGGAAAAATAAAC 19459
RESULT 14
AF340221
LOCUS AF340221 2535 bp DNA linear BCT 22-MAY-2001
DEFINITION Streptococcus pneumoniae PnpA (pnpA) gene, complete cds.
ACCESSION AF340221
VERSION AF340221.1 GI:13447093
KEYWORDS Streptococcus pneumoniae
SOURCE Streptococcus pneumoniae
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 2535)

Db	301	GAAGAATTGCTTGATGAAGGATCCAACTACCGCTAAAGACGAGATATTATCACTGAA	360
QY	101	ValLysGlyGlyTyrValLileLysValAspGlyLysTyrTyrValTyrLeuLysAspAla	120
Db	361	ATCAAGGTTGTTATGTGATTAAAGGTAGATGMAAATACTATGTTTACCTTAAGGATGCA	420
QY	121	AlaHisAlaAspAsnValArgThrLysGluLileAsnArgGlnLysGlnLHisSer	140
Db	421	GCTCATGCGGATAATGTCGATCAAAAGAAGAAATCAATCGACAAAAACAAGCATAGT	480
QY	141	GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln	160
Db	481	CAACATCGTGAAGGTGGGACTCCAAAGAACGATGGTGCTGTTCCTTGGCAAGTTCACAG	540
QY	161	GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr	180
Db	541	GCACGCTACACACAGATGATGTTATATCTTTAATGCTTCTGATATCATTTGAAGATACT	600
QY	181	GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu	200
Db	601	GGTGATGCTTATATCGTTCTCCACGCGCACCATTAACCATTAACATCTAAGAATGAGTTA	660
QY	201	SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn	220
Db	661	TCAGCTACGAGTTAGCTTGCTGAAAAGCCTTCCTATCTGTGTCGGGAAATCTGTCAAA	720
QY	221	SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer	240
Db	721	TCAAGAACCTATGCGGCACAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTCT	780
QY	241	ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln	260
Db	781	GTAAGCAATCCAGAACTACAAATACTAACCAAGCAACACAGCAACTAACAGTCAA	840
QY	261	AlaSerGlnSerAsnAspIleAspSerLeuLysGlnLeuTyrLysLeuProLeuSer	280
Db	841	GCAAGTCAAAGTAATGACATTTGATGTCCTTTGAACACAGCTCTCAAACTGSCCTTTGAGT	900
QY	281	GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr	300
Db	901	CAACGCCATGTAGAATCTGATGGCTTATTTTCACCCAGCGCAAAATCAACAGTCGAACC	960
QY	301	AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet	320
Db	961	GCCAAATGGTGTGTGTGTACCGCAGGAGACCATTAATCACTTATTCCTTATTCACAACTG	1020
QY	321	SerGluLeuGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr	340
Db	1021	TCACCTTTGGAAAGAAATTTGGCTCGTATTATTCCTTCCTTATTCCTTATTCACAACTG	1080
QY	341	ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly	360
Db	1081	GTACCAAGATTCAAGACCAGAACCAACCAAGTCCCAATCGATCCGGAACCTAGTCCAACTG	1140
QY	361	ProGlnProAlaProAsnLeuLys-----IleAspSerAsnSerSer	374
Db	1141	CCGCAACTGCACAAATCTCCAAACAGCTCCCAAGCAATCCAATTTGAT-----GAGAAA	1194
QY	375	LeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIle	394
Db	1195	TTGGTCAAAAGCACTGTTCCAAAAGTAGGCGATGGTTATGCTTTGAGGAGAAATGGAGTT	1254
QY	395	SerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLys	414
Db	1255	CCTCGTTATATCCAGCAAGGATCTTTCAGAGAAACAGCACAGGATTTGATGACAAA	1314
QY	415	LeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaPro	434
Db	1315	CTGGCCAAAGCAGAAAGTTTATCTCAATAGCTAGGAGCTAAGAAAACTGACCTCCCATCT	1374
QY	435	ArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrClnAlaHisLysAlaLeu	454

Db	1375	AGTGTAGAGAAATTTTACAATAAGGCTTATGACTTACTTAGCAAGAATTCATCAAGATT	1434
QY	455	Phe***AsnLysGlyValArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeu	474
Db	1435	CTTGATATAAAGGTCGACAAGTTGATTTGAGGCTTTGGAGTAACCTGTTGGAAACGACTC	1494
QY	475	AsnAspGluSerThrAsnLysGluLysLeuValAspLeuLeuAlaPheLeuAlaPro	494
Db	1495	AAGGATGTCCTCAAGTGATAAGTCAAGTTAGTGGATGATATTCITGCTCTTACGCTCCG	1554
QY	495	IleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGlu	514
Db	1555	AITCGTGTATCCAGAACGTTTAGGAAACCAATCGCAAAATTACTACACTGATGATGAG	1614
QY	515	ValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGlu	534
Db	1615	ATTCAAGTAGCCAAGTTGGCAGGCAGGTACACAACAGAACGGTGTATATCTTTGATCCT	1674
QY	535	HisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHis	554
Db	1675	CGTGATATAACAGTGTAGGGGGATGCCTATGTAACTCCACATATGACCCATAGCCAC	1734
QY	555	TripIleGlyLysAspSerLeuSerAspLysGluLysValIleAlaGlnAlaTyrThrLys	574
Db	1735	TGGATTAANAAGATAGTTTCTCGAAGCTGAGAGCGCAGCCAGGCTTATGCTAAA	1794
QY	575	GluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGlyAsp	594
Db	1795	GAGAAAGGTTTGACCCCTCCTTCACAGACCATCAGGATTCAGAAATATCTGAGGCAAAA	1854
QY	595	SerAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArgLeu	614
Db	1855	GGAGCAAGAGCTATCTACAACCGCGTGAAGCAGCTAAGAAAGGTGCCACTTGATCGTATG	1914
QY	615	ProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHisLys	634
Db	1915	CCTTACAACTCTCAATATCTGTAGAGTCMAAAGCGTAGTTTAATCATCATCTATTAT	1974
QY	635	AspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsn	654
Db	1975	GACCATTTACCATAACATCAATTTGAGTGGTTTGACGAGGCCCTTTATGAGGCACCTAAG	2034
QY	655	GlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGlu	674
Db	2035	GGGTATAGTCTTGAGATCTTTTGGCGACTGTCAAGTACTATGTGGAACATCCAAACGAA	2094
QY	675	ArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAsp	694
Db	2095	CGTCCGCAATTCAGATATGTTTGTGAACGCTAGCGACCATGTTTCAAGAAACAAAAAT	2154
QY	695	HisSerGluAspProAsnLysAsnPheLysAlaAspGluGlu	708
Db	2155	GGTCAAGCTGATACCAATCAAAACGGAAAAACCAACAGAGAGAAACCTCAGACAGAAAA	2214
QY	709	ProValGluGluThr	713
Db	2215	CCTGAGAAAGAAACCCCTCGAGAGAAACCGCAAGACGAAACACAGAGTCTCCAAA	2274
QY	714	-----ProAlaGluProGluValPro	720
Db	2275	CCAACAGAGAACCCAGAGAAATCCACAGAGAAATCACCAGAGAAATCAGAGAAACCT	2334
QY	721	GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys	740
Db	2335	CAGGTCGAGACTGAAAGGTTAAAGAAAACTGAGAGAGGCTGAAGATTTTACTTGGAAAA	2394
QY	741	ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn	760
Db	2395	ATCCAGAAATTCCAATTATCAAGTCCCAATGCCAAAGAGACTCTCACAGGATTAAAAATAAT	2454
QY	761	LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla	780
Db	2455	TTACTATTGGCACCCAGACCAACAAATATATATGGCAGAGCTGAAAACTATTGGCT	2514

Qy	781	LeuLeuLysGlySer	785
Dd	2515	TTATTAAGGAGAGT	2529
RESULT 15			PAT 17-JUL-2003
LOCUS	BD263585	2523 bp	linear
DEFINITION	Novel streptococcus antigens.		
ACCESSION	BD263585		
VERSION	BD263585.1	GI:33073353	
KEYWORDS	JP 2002533123-A/2.		
SOURCE	unidentified		
ORGANISM	unclassified.		
REFERENCE	1 (bases 1 to 2523)		
AUTHORS	Hamel,J., Brodeur,B.R., Pineau,I., Martin,D., Rioux,C. and Charland,N.		
TITLE	Novel streptococcus antigens		
JOURNAL	Patent: JP 2002533123-A 2 08-OCT-2002;		
COMMENT	SHIRE BIOCHEM INC OS S. pneumoniae PN JP 2002533123-A/2 PD 08-OCT-2002 PF 20-DEC-1999 JP 2000591190 PR 23-DEC-1998 US 60/113800 PI JOSSE HAMEL, BERNARD R BRODEUR, ISABELLE PINEAU, DENIS MARTIN, PI CLEMENT RIOUX. PI NATHALIE CHARLAND PC C12N15/09,A61K39/00,A61K39/39,A61P11/00,A61P25/00, PC A61P27/16, PC A61P31/04,C07K14/315,C07K19/00,C12N1/15,C12N1/19,C12N1/21, PC C12N5/10, PC C12P21/02,C12N15/00,C12N5/00 CC Coding region of BVH-11 gene FH key Location/Qualifiers FT CDS (1)..(2520).		
FEATURES	source		
ORIGIN			
Alignment Scores:			
Pred. No.:	1.25e-175	Length:	2523
Score:	3218.00	Matches:	616
Percent Similarity:	82.95%	Conservative:	65
Best Local Similarity:	75.03%	Mismatches:	102
Query Match:	77.26%	Indels:	38
DB:	6	Gaps:	3
US-09-765-271-56 (1-796) x BD263585 (1-2523)			
Qy	1	SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnArgValSerTyr	20
Dd	61	GCTTAGAACTAGTGTTGCATCAGCTCAACTGTAAAGAAATAAATCGTGTTCCTAT	120
Qy	21	IleAspGlyLysGlnAlaThrGlnLysThrGluAenLeuThrProAspGluValSerLys	40
Dd	121	ATAGATGGAAAAACAAGCGACGCAAAAACGGAGAATTTGACCTCGTGAGGTATGCAAG	180
Qy	41	ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr	60
Dd	181	CGTGAAGGAATCAACGCCCAACAAATCGCATCAGATTACGGATCAAGATTATGTACC	240
Qy	61	SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleSer	80
Dd	241	TCTCATGAGAACCATTTACTTACTATATATGCAAGGTCCCCTTATGATGCCATCATCA	300
Qy	81	GluGluLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu	100
Dd	301	GAAGAGCTCTCATGAAGAATCCGAATATVACGTTGAAGATTACAGATTGTCAATGAA	360

QY 455 Phe****AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuGluArgLeu 474
 Db 1435 CTTGTATAATAAAGGTCGACAGTTGATTTTGGGCTTTGGATAACCTGTTGGACGACTC 1494
 QY 475 AsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuAlaPheLeuAlaPro 494
 Db 1495 AAGGATGCTCAAGTGATAAAGTCAAGTTAGTGGATGATATTCTTGCCCTCTTAGCTCG 1554
 QY 495 IleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluThrThrGluAspGlu 514
 Db 1555 ATTGCTCATCCAGAACGTTTAGGAAACCAATAGCGCAATACCTACCTACATGATGAG 1614
 QY 515 ValArgIleAlaGlnLeuAlaAspLysThrThrSerAspGlyTyrIlePheAspGlu 534
 Db 1615 ATTCAAGTAGCCAGTTTGGCAGGCAAGTACACACAGAACGCGTATATCTTTGATCT 1674
 QY 535 HisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHis 554
 Db 1675 CGTATATACCGATGATGAGGGGATGCTGTATGTAACCTCCACATATGACCATAGCCAC 1734
 QY 555 TrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLys 574
 Db 1735 TGGATTAAAGATAGTTTCTCGAAGCTGAGAGCGCGAGCCGCTTATGCTAAA 1794
 QY 575 GluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGlyAsp 594
 Db 1795 GAGAAAGGTTTGCACCTCTTCGACAGACCATCAGGATTCAGGAATACTGAGGCAAAA 1854
 QY 595 SerAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArgLeu 614
 Db 1855 GGACGACAGATCTACTCAACCGCGTGAAGCAGCTAAGAGGTGCCACTTATGCTATG 1914
 QY 615 ProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHisLys 634
 Db 1915 CCTTACAATCTTCAATATCTAGTAGAGTCAAAACGGTAGTTTATCATACCTCATAT 1974
 QY 635 AspHisThrHisAsnIleLysPheAlaTyrPheAspHisThrThrLysAlaProAsn 654
 Db 1975 GACCATTTACCATATCAATCAATTTGAGTGTGTTGACGAGCGCTTTATGAGGCACCTAAG 2034
 QY 655 GlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrThrValGluHisProAspGlu 674
 Db 2035 GGGTATCTCTTGGAGTCTTTGGCCACTGTCAAGTACTATGTGCAACATCCAAACGAA 2094
 QY 675 ArgProHisSerAsnAspGlyTyrPglYAsnAlaSerGlyAlaLeuGlyLysLysAsp 694
 Db 2095 CGTCCGATTTCAGATAATGTTTGGTAAAGCTAGCGACCATGTTTCAAGAAACAAAAAT 2154
 QY 695 HisSerGluAspProAsnLysAsnPheLysAlaAsp----- 706
 Db 2155 GGTCAAGCTGATACCAATCAACCGGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAA 2214
 QY 706 ----- 706
 Db 2215 CCTGAGGAAGAACCCCTCGAAGAGAAACCAACAAAGCGAGAAACAGAGTCTCCAAAA 2274
 QY 707 -----GluGluProValGluGluThrProAlaGluProGluValProGlnValGluThr 724
 Db 2275 CCAACAGAGGAACCAAGAAAGATCAACAGAGGATCAGAGAACCTCAGGTCGAGACT 2334
 QY 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer 744
 Db 2335 GAAAGGTTGAAAGAAACCTGAGAGGCTGAAGATTTACTTGGAAAAATCCAGGATCCA 2394
 QY 745 SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGln 764
 Db 2395 ATTATCAAGTCCATGCCAAAGAGACTCTCAGAGGATTAATAAATTTACTATTGCG 2454
 QY 765 IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGly 784
 Db 2455 ACCCAGGACAAATACTATTATGCGAGAGGCTGAAAAAATCTATTGGCTTTATTAAAGGAG 2514
 QY 785 Ser 785

Db 2515 AGT 2517
 RESULT 16
 AX343072
 LOCUS AX343072
 DEFINITION Sequence 3 from Patent WO0198334.
 ACCESSION AX343072
 VERSION AX343072.1 GI:18152270
 KEYWORDS
 SOURCE unidentifed
 ORGANISM unidentifed
 unclassified.
 REFERENCE 1
 AUTHORS Hamel,J., Ouellet,C., Charland,N., Martin,D. and Brodeur,B.
 TITLE Streptococcus antigens
 JOURNAL Patent: WO 0198334-A 3 27-DEC-2001;
 SHIRE BIOCHEM INC. (CA)
 FEATURES
 Location/Qualifiers
 source
 1..2523
 /organism="unidentifed"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32644"
 /note="BVH-11"
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,25e-175 Length: 2523
 Score: 3218.00 Matches: 616
 Percent Similarity: 82.95% Conservative: 65
 Best Local Similarity: 75.03% Mismatches: 102
 Query Match: 77.26% Indels: 38
 DB: 6 Gaps: 3
 US-09-765-271-56 (1-796) x AX343072 (1-2523)
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
 Db 61 GCTTATGAACGTAGTTTGCATCAAGCTCAAACTGTAAGAAAAATAATCGTTTCTAT 120
 QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
 Db 121 ATAGATGGAAAAAAGCGAGCGAAAAACGAGAAATTTGACTCTGATGAGTTAGCAAG 180
 QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
 Db 181 CGTGAAGGAATCAACGCCGACAAATCGTCATCAAGATTACGGATCAAGTTTATGTGACC 240
 QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
 Db 241 TCTCATGGAGACCATTTATCTACTATAATGCAAGGTCCCTTATGATGCCATCATCAGT 300
 QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
 Db 301 GAAGAGCTCTCATGAAGATCCGAATATCATGTTAGGATTCAGACATTGTCAATGAA 360
 QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
 Db 361 ATCAAGGGTGGTTATGTCTTAAGGTAAACGGTAAATACTAIGTTTACCTTAAAGGATGA 420
 QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
 Db 421 GCTCATGGGATTAATGTCGTACAAAAGAGAAATCAATCGCAAAAAACAAGACATAGT 480
 QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
 Db 481 CAGCATCTGAGGAGGAGGACTTCAGCAACCGATGCTGGGTAGCCCTTGCACGTTACAG 540
 QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
 Db 541 GGACGCTACACACAGATGATGGTTATATCTTCAATGATCTGATCATCGAAGATACG 600
 QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200

Novel streptococcus antigens
Patent: JP 200253123-A 7 08-OCT-2002;

QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
 Db 945 CAACGCCATGTAGAATCTGATGCTTATTTTCGACCCAGCGCAATCACAGTGAACC 1004
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
 Db 1005 GCCAGAGTGTAGCTGCTCCCTCATGTAACCAATACCACTTTATCCCTTATGAACAAATG 1064
 QY 321 SerGluLeuGluGluArgIleAlaArgIlePheProLeuArgTyrArgSerAsnHisTrp 340
 Db 1065 TCTGAATTGGAAACAAAGTCTCTGTAATATCCCTTCCTTCGTTATCGTTCAACCAATGG 1124
 QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
 Db 1125 GTACCAAGTTCACAGACCAAGAACCAAGTCCACAAACGACTCCAGAACCTAGTCCAAAGT 1184
 QY 361 ProGlnProAlaProAsnLeuLys-----IleAspSerAsnSerSer 374
 Db 1185 CCGCAACTCTGACCAAAATCTCAACAGCTCCAAAGCAATCAATATGAT-----GAGAAA 1238
 QY 375 LeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIle 394
 Db 1239 TTGCTCAAGAGAGCTGTTCGAAGAGTAGGCGATGCTTAAGTCTTTGAGGAGAAATGGAGTT 1298
 QY 395 SerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLys 414
 Db 1299 TCTCGTTATATCCAGGCAAGAACTCTTCAGCAGAAACAGCAGCAGCAATGATAGCAAA 1358
 QY 415 LeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaPro 434
 Db 1359 CTGGCCAAAGCAGGAAGTGTATCTCAAGCTAGGAGCTAAGAAACTGACCTCCCATCT 1418
 QY 435 ArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeu 454
 Db 1419 AGTGATCGAGATTTTACAATAGGCTTATGACTTACTAGCAAGAAATCCACCAAGATTTA 1478
 QY 455 Phe***AsnLysGlyArgAsnSerAspPheAlaLeuAspLysLeuLeuGluArgLeu 474
 Db 1479 CTTGATAATAAAGGTCACAAAGTTGATTTTGAGGCTTTGGATAACCTTTGGACGACTC 1538
 QY 475 AsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaPro 494
 Db 1539 AAGATGCTCAAGTGATAAAGTCAAGTTAGTGTGATGATATCTTGTCTTCTTAGCTCCG 1598
 QY 495 IleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGlu 514
 Db 1599 ATTGCTCATCCAGACGTTTAGAANAACCAATGCGCAAAATACCTACACTGATGATGAG 1658
 QY 515 ValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGlu 534
 Db 1659 ATCAAGTAGCCCAAGTTGGCAGGCAAGTACACAAACAGACGCTTATATCTTTGATCCT 1718
 QY 535 HisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHis 554
 Db 1719 CGTGATATAACAGTATGAGGGGATGCTATGTAACTCCACATATGACCCATAGCCAC 1778
 QY 555 TrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLys 574
 Db 1779 TGGATTAAAAAGATAGTTTGTGAGCTGAGAGCGCAGCGCAGCCGCTTATGCTAAA 1838
 QY 575 GluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGlyAsp 594
 Db 1839 GAGAAGGTTTACCCCTCTCCACAGACCATCAGGATTCAGGAATATCTGAGCAAAA 1898
 QY 595 SerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArgLeu 614
 Db 1899 GGAGCAGAAGCTATCTACAAACCGGTGAAGCAGCTAAGAGAGTGCACCTGATCGTAGT 1958
 QY 615 ProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHisLys 634
 Db 1959 CTTTACAATCTTCAATATCTAGTAGTCAAAACCGGTAGTTTAAATCATACCTCATTTAT 2018
 QY 635 AspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsn 654

Db 2019 GACCATTACCAATAACATCAAAATTTGAGTGGTTGACGAAGGCTTTATGAGCACCTAAG 2078
 QY 655 GlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGlu 674
 Db 2079 GGTATATCTTTGAGATCTTTTGGCACTGTGCAAGTACTATGTGCAACATCCAACGAA 2138
 QY 675 ArgProHisSerAsnAspGlyTyrPglyAsnAlaSerGluHisValLeuGlyLysLysAsp 694
 Db 2139 CGTCGCAATTCAGATAATGTTTGTAAACGCTGACGACCATGTTCAAGAACAACAAAAT 2198
 QY 695 HisSerGluAspProAsnLysAsnPhelysAlaAsp----- 706
 Db 2199 GGTCAAGCTGATACCAATCAAAACGAAAAACCAAGCAGGAGAAACCTCAGACAGAAA 2258
 QY 706 ----- 706
 Db 2259 CCTGAGGAGAAACCCCTCGAGAGAGAGAAACCAAAACGAGAACCCAGAGTCTCCAAA 2318
 QY 707 -----GluGluProValGluGluThrProAlaGluProGluValProGluValGluThr 724
 Db 2319 CCAACAGAGGAAACCAAGAGAAATCACAGAGGAATCAGAGAACCTCAGTCCAGACT 2378
 QY 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer 744
 Db 2379 GAAAGGTTGAAGAAAAAAGTCTGAGAGAGGCTGAAGATTTACTTGGAAAAATCCAGATCCA 2438
 QY 745 SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnLeuThrLeuGln 764
 Db 2439 ATTATCAAGTCCAAATGCAAGAGACTCTCAGAGATTAAAAATAATTTACTATTGGC 2498
 QY 765 IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGly 784
 Db 2499 ACCAGGACCAACAATACTATTATGGCAGAGAGCTGAAAAAATTTGGCTTTTAAAGAG 2558
 QY 785 Ser 785
 Db 2559 AGT 2561
 RESULT 19
 BD229972
 LOCUS
 DEFINITION Human complement C3-degrading polypeptides from streptococcus pneumoniae.
 ACCESSION BD229972
 VERSION BD229972.1 GI:33039742
 KEYWORDS JP 2002526082-A/3
 SOURCE Streptococcus pneumoniae
 ORGANISM Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 REFERENCE 1 (bases 1 to 2478)
 AUTHORS Hostetter,M.K., Finkel,D.J., Cheng,Q., Green,B.A. and Masi,A.W.
 TITLE Human complement C3-degrading polypeptides from streptococcus
 JOURNAL Patent: JP 2002526082-A 3 20-AUG-2002;
 REGENTS OF THE UNIVERSITY OF MINNESOTA,AMERICAN CYANAMID CO
 COMMENT OS Streptococcus pneumoniae
 PN JP 2002526082-A/3
 PD 20-AUG-2002
 PF 24-SEP-1999 JP 2000574269
 PR 24-SEP-1998 US 60/101736,31-MAR-1999 US 09/283094 PI
 MARGARET K HOSTETTER, DAVID J FINKEL, QI CHENG, BRUCE A GREEN, AMY
 PI W MASI
 PC C12N15/09,A61K38/00,A61K39/09,A61K39/395,A61K39/395,A61K48/00,
 PC A61P29/00,
 PC A61P37/04,A61P37/06,A61P43/00,C07K16/40,C12N1/15,C12N1/19, PC
 C12N1/21,
 PC C12N5/10,C12N9/48,C12P21/08,C12N15/00,C12N5/00,A61K37/02 CC
 Human complement C3-degrading polypeptides from streptococcus CC
 pneumoniae
 FH Key Location/Qualifiers
 FT source l. .2478
 FT /organism='Streptococcus pneumoniae'.

FEATURES

Location/Qualifiers
1. .2478
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/db_xref="taxon:1313"

ORIGIN

Alignment Scores:
Pred. No.: 2,35e-152 Length: 2478
Score: 2812.00 Matches: 552
Percent Similarity: 77.10% Conservative: 81
Best Local Similarity: 67.24% Mismatches: 136
Query Match: 67.52% Indels: 52
DB: 6 Gaps: 8

US-09-765-271-56 (1-796) x BD229972 (1-2478)

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAspArgValSer 19
Db 61 TCCTATGAACCTGGTCTGTCACCAAGCTGGTCTGAGTTAAGAAAGAGTCTAATCGAGTTTCT 120
Qy 20 TyrlleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
Db 121 TATATAGATGGTATGATGAGCTGGTCAAAAGGCAGAAACTTGACACCAAGATGAAGTCAGT 180
Qy 40 LysArgGluGlyLysGlnAlaGluGlnLeuValleLysLysLysLysLysLysLysLysLysLys 59
Db 181 AAGAGGAGGGGATCAACGCCGCAACAAATCGTCATCAAGATTAACGATCAAGGTTATGTG 240
Qy 60 ThrSerHisGlyAspHisThrHisTyrTyrAsnGlyLysValProTyrAspAlaAlaIle 79
Db 241 ACCTCTCATGGAGACCAATATCATTAATGCAAGGTCCTTATGATGCCATCATC 300
Qy 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspLysLysVal 99
Db 301 AGTGAAGAGCTCTCATGAAGATCCGAATATCATGTTCAAGGATTCAGACATTCGTAAT 360
Qy 100 GluValLysGlyGlyTyrValleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
Db 361 GAAATCAAGGGTGGTTATGTTATCAAGGTAGATGGAAATATCTATGTTTACCTTAAGGAT 420
Qy 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluLysLysLysLysLysLysLysLys 139
Db 421 CGAGCTCATCGCGGATTAATTCGACAAAGAGAGATTAACGTCAGAGCAGGAACAC 480
Qy 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
Db 481 AGTCATAATCACGGGGTGGTTCT---AACGATCAAGCAGTAGTTGCAGCCAGAGCC 534
Qy 160 GlnGlyArgTyrThrThrAspAspGlyTyrLysPheAsnAlaSerAspLysLysLysLys 179
Db 535 CAAGGACGGCTATACACGGATGATGTTATCTTCAATGCATCTCATATCATTTAGGAC 594
Qy 180 ThrGlyAspAlaTyrLysValProHisGlyAspHisTyrHisTyrLysProLysAsnGlu 199
Db 595 ACGGGTGATCTTATATCGTTCTCTACGGCAGCATTACATTTACATTCCTAAGATGAG 654
Qy 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
Db 655 TTATCAGTAGCAGTGTAGTGTGTCAGAGGCTTATTTGGAATGGG----- 699
Qy 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239
Db 700 -----AAGCAGGGATCTCGTCTTCCTTCATGTTCTAGTTATATGCA 741
Qy 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
Db 742 AATCCAGCTCAACCAAGATTGTCAGAGACCAACATCTGACTGTCATCTCAACTTATCAT 801
Qy 260 GlnAlaSerGlnSerAsnAspLysAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
Db 802 CAA---AATCAAGGGGAAACATTTACAGCCTTTTACGTGAATTTGATGCTAAACCCCTTA 859

Qy 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
Db 859 TCAGAACGCCATCTGGAATCTGATGGCTTATTTTCAGCCAGCGCAATACCAAGTCGA 918
Qy 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
Db 919 ACCGCCAGAGGTGATGCTGCTCTCATGGTAACCAATACCATTTATCCCTTATGAACA 978
Qy 320 MetSerGluLeuGluGluArgIleAlaArgIleLeuProLeuArgTyrArgSerAsnHis 339
Db 979 ATGCTCGAATGGAAACCAAGATTCGTCGTATTATTCCTCTCGTTATCGTTCAACCAT 1038
Qy 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
Db 1039 TGGGTACAGATTCAAGACCAGAACCAACCAAGTCCCAATCGACTCCGGAACCTTAGTCCA 1098
Qy 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373
Db 1099 AGTCGGCAACCTGCACCAATCTCAACAGCTCCAAGCAATCCCAATTTGAT-----GAG 1152
Qy 374 SerLeuValSerGlnLeuValArgLysValGlyGlyTyrValPheGluGluLysGly 393
Db 1153 AAATTTGGTCAAGAGAGCTGTTTCGAAAGTAGTGGCGATGGTTATGCTTTTGGAGAGAATGGA 1212
Qy 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413
Db 1213 GTTCTCTGTTATATCCCAAGCAAGATCTTTTCAGCAAGAACAGCAGCAGGAGCATTTGATAGC 1272
Qy 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433
Db 1273 AAATGGCCAGCAGAGAAAGTTTATCTCATAGCTTAGGAGCTAAGAAACCTGACCTCCA 1332
Qy 434 ProArgAspGlnGluPheTyrAsnLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453
Db 1333 TCTAGTGTATCGAATTTTACAATTAAGCTTATGACTTACTAGCAAGAATTCACCAAGAT 1392
Qy 454 LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473
Db 1393 TTACTTTGATAATAAAGTTCGACAGTGTATTTTGAGGCTTTGGATAACCTGTTTGAACGA 1452
Qy 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuAlaPheLeuAla 493
Db 1453 CTCAGGATGTCCTCAAGTATTAAGTCAAGTTAGTGGATGATATTTCTTCCCTTCTTAGCT 1512
Qy 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513
Db 1513 CCGATTCGTATCCAGACGTTTTCAGAAACCAATTCGCAAAATACCTACATCATGATGAT 1572
Qy 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533
Db 1573 GAGATTCAGTAGCCCAAGTTGGCAGCAAGTACACCAACAGACAGCGTTATATCTTTGAT 1632
Qy 534 GluHisAspIleLeuSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553
Db 1633 CCTCGTATATACAGTGTATGAGGGGATGCTATGTAATTCACATATGACCCATAGC 1692
Qy 554 HisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr 573
Db 1693 CACTGGATTAATAAAGATAGTTTCTGTAAGCTGAGAGAGCGCAGCCAGGCTTATGCT 1752
Qy 574 LysGluLysGlyLysLeuProSerProAspAlaAspValLysAlaAlaProThrGly 593
Db 1753 AAAGAGAAAGGTTTTCAGCTCTCTGACAGACCATCAGGATTCAGGAATATCTAGGCA 1812
Qy 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArg 613
Db 1813 AAAGGACGAAAGCTATCTACACCGCTGAAAGCAGCTAAGAGGTCGCCATTCATTCGT 1872
Qy 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633
Db 1873 ATGCTTCAATCTTCAATATCTGTAAGAGTCAAAACCGGTAGTTTAAATCATACCTCAT 1932
Qy 634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653

```

Db      1933 TATGACCATTTACCAATCAAAATTTGAGTGGTTTGACGAAGCCCTTTATGAGGACCT 1992
QY      654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673
Db      1993 AAGGGGTATACTCTGAGGATCTTTTGGCGACTGTCAAGTACTATGCGAACATCCAAAC 2052
QY      674 GluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLys 693
Db      2053 GAACGTCGCGCATTCAGATAATGTTTGTGTAAACGCTAGCGACCATGTTCAAGAAACAAA 2112
QY      694 AspHisSerGluAspProAsnLysAsnPheLysAlaAspGluGlu 708
Db      2113 AATGTCAGCTGATACCAATCAACGGAACCAAGCGAGAGAAACCTCAGACAGAA 2172
QY      709 ---ProValGluGluThr 713
Db      2173 AAACCTGAGGAAGAAACCCCTCGAGNAGAGAAACCGCAAGGAGAGAAACCGAGTCTCCA 2232
QY      714 -----ProAlaGluProGluValProGluValProGluValGluThr 724
Db      2233 AAACCAACAGAGAAACCAAGAAATCACAGAGGAATCAGAAAGAACCTCAGTCCAGACT 2292
QY      725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer 744
Db      2293 GAAAGGTTTGAAGAAAACTGAGAGAGCTGAAGATTTTACTTGGAAAAATCCAGGATCCA 2352
QY      745 SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuAlaArgAsnAsnLeuThrLeuGln 764
Db      2353 ATTATCAAGTCCATGCAAGAGACTCTCAGAGATTAAAAATAATTTACTATTGGC 2412
QY      765 IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLeuLysGly 784
Db      2413 ACCAGGACAACTACTATTATGCGAGAACTGAAAACTATTGGCTTTATTAAAGAG 2472
QY      785 Ser 785
Db      2473 AGT 2475

RESULT 20
AX569137
LOCUS      2457 bp DNA linear PAT 29-NOV-2002
DEFINITION Sequence 2345 from Patent WO207021.
ACCESSION AX569137
VERSION    AX569137.1 GI:26002635
KEYWORDS   Streptococcus pneumoniae
SOURCE      Streptococcus pneumoniae
ORGANISM    Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE   1
AUTHORS     Masignani, V., Tettelin, H. and Fraser, C.
TITLE       Streptococcus pneumoniae proteins and nucleic acids
JOURNAL     Patent: WO 0207021-A 2345 03-OCT-2002;
            Chiron Spa (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
            Location/Qualifiers
            1..2457
               /organism="Streptococcus pneumoniae"
               /mol_type="unassigned DNA"
               /db_xref="taxon:1313"
ORIGIN
Alignment Scores:
Pred. No.:      1.58e-151      Length:      2457
Score:          2797.50      Matches:      546
Percent Similarity: 77.30%      Conservative: 84
Best Local Similarity: 66.92%      Mismatches: 138
Query Match:     67.11%      Indels:      47
Db:              6              Gaps:      7

US-09-765-271-56 (1-796) x AX569137 (1-2457)
QY      1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnAsnArgValSer 19

```

```

Db      61 TCCTATGAGCTTGAGCGTTACCAAGCTGGTCAGGATAAGAAAGAGTCTAATCGAGTTGCT 120
QY      20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
Db      121 TATATAGATGGTGATCAGGCTGGTCAAAAGGCGAGAAACTTCACACAGATGAAGTCAGT 180
QY      40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
Db      181 AAGAGGAGGGGATCAACGCCCAACAAATGTTATCAAGATTACGGATCAAGGTTATGTG 240
QY      60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79
Db      241 ACCTCTCATGGAGACCAATTATCTATATAATGCAAGGTTCTTATGATGCCATCATC 300
QY      80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
Db      301 AGTGAAGAGCTCTCATGAAGATCCGAATATATCAGTTGAAGGATTCAGACATTGTCAAT 360
QY      100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
Db      361 GAAATCAAGGTTGGTTATGTCTAATAGGTAACGGTAATACTATGTTTACTCTTAAGGAT 420
QY      120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
Db      421 GCAGCTCATGCGGATAATATTCGACAAAAGAGAGATTAAACGTCAGACAGAGNACGC 480
QY      140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
Db      481 AGTCATATCATAACTCA-----AGAGCAGATAATGCTGTGTCAGCCAGAGACC 531
QY      160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
Db      532 CAAGGAGCTTATACACGAGATGCGGTATATCTTCAATGCATCTGATATCATTCATTGAGAC 591
QY      180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
Db      592 ACGGTTGATGCTTATATCGTTCTCTCACGGCGACCATTCACATTACATTCCTTAAGATGAG 651
QY      200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
Db      652 TTATCAGCTAGCGAGTTAGTCTGTCAGAAAGCCTATTGGAATGGG-----696
QY      220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239
Db      697 -----AAGCAGGAGATCTCGCTCTCTTCAAGTTCTTCAAGTTCTTAATATGCA 738
QY      240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
Db      739 AATCCAGCTCAACCAAGATTGTCAGAGAACCAATCTGACTGTCTCACTCCAACCTTATCAT 798
QY      260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
Db      799 CAA---AATCAAGGGGAAACCAATTTCAAGCCTTTTACGTGAATTTGATGCTAAACCTTA 855
QY      280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
Db      856 TCAGAACGCCATGTGGAATCTGATGCGCTTATTTTCGACCCCGCAACAAATCAAGTCCA 915
QY      300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
Db      916 ACCGCCAGAGGTGAGTGTGCTCATGTAACCATTAACCATTTATCCCTTATGCAACA 975
QY      320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
Db      976 ATGTCGTAATTTGGAACAAACGAATTTGCTGTTATTTCCCTTCGTTATCGTTCAAAACCAT 1035
QY      340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
Db      1036 TGGGTACCAGATTCAAGACCAAGAGAACCAAGTCCACAGACCTCCAGAACCTAGTCCA 1095
QY      360 GlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeu 379

```


1096 AGTCGCAACAGCTCCAGCAATCAATGAT-----GAGAAATGGTCAAGAAGCT 1149
 Db
 380 ValatcLysValGlyGluGlyTyrValPheGluGluGlyGlyIleSerArgTyrValPhe 399
 QY
 1150 GTTCGAAAGTAGCGATGTTATGCTTTGAGGAGAAATGGAGTCTCTGTTATATCCCA 1209
 Db
 400 AlalyAspLeuProSerGluThrValLysAsnLeuGluSerLysLysGluGlu 419
 QY
 1210 GCCAAGATCTTCAGCAGAAACAGCAGCAGCAATGATAGCAAACTGGCCAAAGCAGAA 1269
 Db
 420 SerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGluGluPhe 439
 QY
 1270 AGTTATCTCATAGCTAGGAACTAAGAACTGACCTCCCATCTAGTGTGATCGAGATTT 1329
 Db
 440 TyrAspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGly 459
 QY
 1330 TACAATAGGCTTAAGCTTACTAGCAAGATTCACCAAGATTTTACTGATAATAAAGGT 1389
 Db
 460 ArgAsnSerAspPheGlnAlaLeuAspLysLeuGluArgLeuAsnAspGluSerThr 479
 QY
 1390 CGACAAGTTGATTTAGGCTTTGATTAACCTGTTGGAACGACTCAAGATGTTCTCAAGT 1449
 Db
 480 AsnLysGluLysLeuValAspLeuAlaPheLeuAlaProIleThrHisProGlu 499
 QY
 1450 GATAAAGTCAAGTTAGTGAAGATATCTTGCTTCTTAGCTCCGATTCGTCAATCCAGAA 1509
 Db
 500 ArgLeuGlyLysProAsnSerGlnIleGluThrGluAspGluValArgIleAlaGln 519
 QY
 1510 GCTTTAGGAAACCAATGCGCAATTACTACACTGATGATGATGATCAAGTAGCCAAAG 1569
 Db
 520 LeuAlaAspLysTyrThrSerAspGlyTyrIlePheAspGluHisAspIleSer 539
 QY
 1570 TTGGCAGGCAAGTACACACAGAACGCGTTATATCTTTGATCCTCGTATATACCAAGT 1629
 Db
 540 AspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisIleGlyLysAsp 559
 QY
 1630 GATGAGGGGGATGCTTATGATCTCCATATATGACCATGACCATGATTAATAAAGAT 1689
 Db
 560 SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeu 579
 QY
 1690 AGTTTGTCTGAAGTGAAGAGCGGCGAGCCAGCTTATGCTTAAGAGAAAGGTTTGACC 1749
 Db
 580 ProProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIle 599
 QY
 1750 CTTCTCTTCAGACACCATCAGGATTCAGGAAATCTGAGCAGAAAGAGCAGAGCTATC 1809
 Db
 600 TyrAsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGlu 619
 QY
 1810 TACAACCGGTGAAGACGCTAAGAGGTGCGCATCTGATGCTTACATCTTCAATCTTCAA 1869
 Db
 620 HisThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsn 639
 QY
 1870 TATCTGTAGAGTCAAAACGCTAGTTTAAATCATCTCATCTATGACCATTAACATAAC 1929
 Db
 640 IleLysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGlu 659
 QY
 1930 ATCAAAATTTGAGTGGTTTGAAGAGCGCTTTATGAGGCACCTAAGGGGTATCTCTTCA 1989
 Db
 660 AspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsn 679
 QY
 1990 GATCTTTTGGCGACTGTCAGTACTATGTGCAACATCCAAACGAACTGCGCATTCAGAT 2049
 Db
 680 AspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspPro 699
 QY
 2050 AATGGTTTGGTAACGCTAGCAGCATGTTTCAAGAGAAACAAATAATGGTCAAGCTGATAC 2109
 Db
 700 AsnLysAsnPheLysAlaAspGluGlu-----ProValGluGluThr 713
 QY
 2110 AATCAACGGAACCAACGAGGAGAAACCTCAGCAGAAACCACTCAGAGAAACCA 2169
 Db
 713 ----- 713
 Db
 2170 CCTCAGAGAGAAACCGCAAGCGAGAACCAAGAGTCTCCAAACCAACCAAGAGAACCA 2229

714 -----ProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGln 730
 QY
 2230 GAAGATCACCAGAGGAATCAGAAAGAACTCAGTCCAGACTGAAAGGTTTGAAGAAA 2289
 Db
 731 LeuLysGluAlaGluValLeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAla 750
 QY
 2290 CTGAGAGAGGCTGAAGATTTACTTGGAAAAATCCAGATCCCAATTATCAAGTCCAATGCC 2349
 Db
 751 ThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnSer 770
 QY
 2350 AAGAGAGCTCTCAGAGATTTAAAAATAATTTACTATTGGCACCAGGACACAAATACT 2409
 Db
 771 IleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer 785
 QY
 2410 ATTATGCGAGAGCTGAAAACTATTGGCTTTATTAAGAGAGT 2454
 Db
 RESULT 21
 BD264091 2481 bp DNA linear PAT 17-JUL-2003
 LOCUS Streptococcus pneumoniae proteins and nucleic acid molecules.
 DEFINITION
 BD264091
 ACCESSION BD264091.1 GI:33073859
 VERSION JP 2002531055-A/27.
 KEYWORDS Streptococcus pneumoniae
 SOURCE Streptococcus pneumoniae
 ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 REFERENCE 1 (bases 1 to 2481)
 AUTHORS Gilbert,C.F.G. and Hansbro,P.M.
 TITLE Streptococcus pneumoniae proteins and nucleic acid molecules
 JOURNAL Patent: JP 2002531055-A 27 24-SEP-2002;
 MICROBIAL TECHNIQS LTD
 COMMENT OS Streptococcus pneumoniae
 PN JP 2002531055-A/27
 PD 24-SEP-2002
 PF 27-JUL-1999 JP 2000562519
 PR 27-JUL-1998 GB 9816337.1,19-MAR-1999 US 60/125164 PI
 CHRISTOPHE FRANCOIS GUY GILBERT,PHILIP MICHAEL HANSBRO PC
 C12N15/09,C12N15/09,A61K39/09,A61K45/00,A61K48/00,A61P31/04, PC
 C07K14/315.
 PC C07K16/12,C12Q1/14,C12Q1/68,G01N33/53,G01N33/566,G01N33/569//
 PC C12P21/08,C12Q1/15,G01N33/50,(C12N15/09,C12R1:46),(C12Q1/14,C12R1:46),
 PC C12N15/00,
 PC C12N15/00,(C12N15/00,C12R1:46)
 CC Streptococcus pneumoniae proteins and nucleic acid molecules
 FH key Location/Qualifiers
 FT source 1..2481
 FT /organism='Streptococcus pneumoniae'.
 FEATURES
 source
 1..2481
 /organism='Streptococcus pneumoniae'
 /mol_type='genomic DNA'
 /db_xref='taxon:1313'
 ORIGIN
 Alignment Scores: 2,22e-151 Length: 2481
 Pred. No.: 2795.00 Matches: 549
 Score: 76.98% Conservative: 83
 Percent Similarity: 66.87% Mismatches: 137
 Best Local Similarity: 67.11% Indels: 52
 Query Match: 6 Gaps: 8
 DB:
 US-09-765-271-56 (1-796) x BD264091 (1-2481)
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnArgValSer 19
 Db 61 TCCTATGAGCTTGTGCTGCTCAACAGCTGGTCAAGATGAGATGAGAGTCTATCGAGTTGCT 120
 QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39

121 TATATAGATGGTGCATCAGGCTGGTCAAAAGCGAGCAAAACTTGACACCATGAAGTCAGT 180
QY
40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
Db
181 AAGAGGGAGGGGATCAACCGCCGACAAATCGTCATCAAGATTACGGATCAAGGTTATGTG 240
QY
60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79
Db
241 ACCCTCATGGAGACCAATTATCATTTACTATATATGCAAGGTCCTTATGATGCCATCATC 300
QY
80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
Db
301 AGTGAAGAGCTCCTCATGAAGATCCGAATATATCATAGTTGAAGGATTCAGACATTGTCAAT 360
QY
100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
Db
361 GAAATCAAGGGTGGTTATGTCTATCAAGGTAGACGGAAATATCTATGTTTACCTTAAGGAT 420
QY
120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
Db
421 GCAGCTCATGGGATATATTCGGACAAAGAGAGATTAAGCGTCAGAAAGCAAGCAGCC 480
QY
140 SerGlnHisAspGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
Db
481 AGTCATATATCAC-----GGGTCAAGAGCTAACCATCATGCAGTAGCTGCAGCAGAGCC 534
QY
160 GlnGlyArgTyrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
Db
535 CAAGGACGCTATACAACGGATGATGGGTATATCTTCAATGATCTGATATCATTTAGGAGC 594
QY
180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
Db
595 ACGGGTCATGCTTATATCGTTCCTCACGGCAGCATTACCATTAACATTCCTAAGAAATGAG 654
QY
200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
Db
655 TTATCAGTACGAGGTAGCTGCTGCGAAGCCCTATTGGAATGGG----- 699
QY
220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239
Db
700 -----AAGCAGGATCTCGTCTCTTCTCAAGTCTCTAGTTCTATGTTAATGCA 741
QY
240 SerValSerAsnProGlyThrThrAsnThrSerSerAsnAsnSerAsnThrAsnSer 259
Db
742 AATCCAGCTCAACCAAGATTGTGAGAAACCAACATCTGACTGTCACTCCCACTTATCAT 801
QY
260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
Db
802 CAA--AATCAAGGGGAACATTTCAAGCCTTTATAGTGAATGTATGCTAAACCCCTTA 858
QY
280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
Db
859 TCAGAACCGCATGTGAAATCTGATGCGCTTATTTTCGACCCAGCGCAATATCAAGTCGA 918
QY
300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
Db
919 ACCGCCAGAGGTAGTACGTGCTCCATGTTGTAACCATTAACCTTTATCCCTTATGAACA 978
QY
320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
Db
979 ATGCTGTAATTGGAAAGAAATGCTGCTATATTTCCCTCTGTTATGCTTCAAAACCAT 1038
QY
340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
Db
1039 TGGGTACCAAGATTCAAGACCAGACACCAAGTCCCAATTCGACTCCGGAACCTAGTCCA 1098
QY
360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373
Db
1099 AGTCGCGACCTGCGACCAATCTCAACCGAGCTCCAAAGCAATCCAAATGAT-----GAG 1152
QY
374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGlyLysGly 393
Db
1153 AAATTGGTCAAGAAAGCTGTTGCAAAAGTAGGCGATGTTATGTTCTTTGAGGAGAAATGCA 1212

QY 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413
Db 1213 GTTCTCGTTATATCCAGCCCAAGGATCTTTTCAGCAGAAACAGCAGCAGCGATTGATAGC 1272
QY 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433
Db 1273 AAACCTGGCCACAGCAGGAAAGTTTATCTCATAGCTAGGAGCTAAGAAAACCTGACCTCCA 1332
QY 434 ProArgAspGlnGluPheTyrAspLysValAsnLeuLeuThrGluAlaHisLysAla 453
Db 1333 TCTAGTGATCGAAGATTTTACAATAAGGCTTATGATGCTTACTAGCAAGAAATTCACCAAG 1392
QY 454 LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473
Db 1393 TTACTTGATATAATAAGGTCGACAAAGTTGATTTGAGGCTTTGGATAACTGTTGAAACGA 1452
QY 474 LeuAsnAspGluSerThrAsnLysGlnLysLeuValAspAspLeuLeuAlaPheLeuAla 493
Db 1453 CTCAGGATGTCCCAAGTGTATAAAGTCAAGTTAGTGGATGATATTTCTTGCCTTCTTAGCT 1512
QY 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513
Db 1513 CCGATTCGTCAATCCAGAACGTTTAGGAAAACCAATGCGCAATTTACCTACACTGATGAT 1572
QY 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533
Db 1573 GAGATTCAGTAGGCAAGTTGGCAGGCAAGTACACACAGAGACGGTTATATCTTTGAT 1632
QY 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553
Db 1633 CCTCGTATATAACCACTGATGAGGGGATGCTTATGTAATCCACATATGACCCATAGC 1692
QY 554 HistripIleGlyLysAspSerLeuSerAspLysGlnLysValAlaAlaGlnAlaTyrThr 573
Db 1693 CACTGGATTAAAAAAGATAGTTTGTCTGAAGCTGAGAGCGGCGAGCCAGGCTTATGCT 1752
QY 574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593
Db 1753 AAAGAGAAAGTTTGAACCCCTCTCGACAGACCATCAGGATTCAGGAATATCTGAGGCA 1812
QY 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyGlnLysArgIleProLeuValArg 613
Db 1813 AAGAGGACGAGAGCTATCTACAAACCGCTGAAAGCAGCTAAGAAAGGTGCACCTGATCGT 1872
QY 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633
Db 1873 ATGCTTACAACTTCAATATCTAGTAAGTCAAAACCGTAGTATTAATCATACCTCAT 1932
QY 634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653
Db 1933 TATGACCATTACATAACATCAAAATTTGAGTGGTTTGACGAAGGCCCTTTATGAGGACCT 1992
QY 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673
Db 1993 AAGGGGTATACTCTTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAAC 2052
QY 674 GluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLys 693
Db 2053 GAACTCGCATTCAGATAATGTTTGTGTAACGCTAGCAGCCTGTTTCCAAAGAACAAA 2112
QY 694 AspHisSerGluAspProAsnLysAsnPheLysAlaAspGluGlu----- 708
Db 2113 AATGCTCAAGCTGTATACCAATCAACAGGAAACCAACGAGGAGGAAACCTCAGACAGAA 2172
QY 709 ---ProValGluGluThr----- 713
Db 2173 AAACCTGAGGAAGAAACCCCTCGAGAGAGAAACCGCAAAAGCGAGAAACACGAGTCTCCA 2232
QY 714 -----ProAlaGluProGluValProGlnValGluThr 724
Db 2233 AAACCAACAGAGGAACCAAGAAATTCACGAGGAATCAGAAAGAACCTCAGGTCGAGACT 2292

QY	725	GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer	744	
Db	2293	GAAGAAGTTGAGAGAACTCAGAGAGCTGAGAGTTTACTTGGAAAAATCCAGATCCA	2352	
QY	745	SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGln	764	
Db	2353	ATTATCAATGCAATGCGCAAGAGACTCTCACAGATTAAAAATAATTTACTATTGGC	2412	
QY	765	IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLeuLysGly	784	
Db	2413	ACCAGGACACAAATCTATTATGTCAGAGCTGAAAACTATTGCTTTTATAAGGAG	2472	
QY	785	AGT	785	
Db	2473	AGT	2475	
RESULT	22			
LOCUS	AF318954	2457 bp DNA linear	BCT 11-FEB-2001	
DEFINITION	Streptococcus pneumoniae pneumococcal histidine triad protein B precursor (phtB) gene, partial cds.			
ACCESSION	AF318954			
VERSION	AF318954.1	GI:12744741		
KEYWORDS	Streptococcus pneumoniae			
SOURCE	Streptococcus pneumoniae			
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.			
REFERENCE	1 (bases 1 to 2457)			
AUTHORS	Adamou, J.E., Heinrichs, J.H., Erwin, A.L., Walsh, W., Gayle, T., Dormitzer, M., Dagan, R., Brewah, Y.A., Barren, P., Lathigra, R., Langermann, S., Koenig, S., and Johnson, S.			
TITLE	Identification and characterization of a novel family of pneumococcal proteins that are protective against sepsis			
JOURNAL	Infect. Immun. 69 (2), 949-958 (2001)			
MEDLINE	21101045			
PUBMED	11159990			
REFERENCE	2 (bases 1 to 2457)			
AUTHORS	Adamou, J.E., Heinrichs, J.H., Erwin, A.L., Walsh, W., Dormitzer, M. and Johnson, S.			
TITLE	Direct Submission			
JOURNAL	Submitted (03-NOV-2000) Molecular Microbiology, MedImmune, Inc., 35 West Watkins Mill Road, Gaithersburg, MD 20878, USA			
FEATURES	source			
gene	1..2457	/organism="Streptococcus pneumoniae"		
CDS	1..2457	/mol_type="genomic DNA"		
	1..2457	/serotype="4"		
	1..2457	/db_xref="taxon:1313"		
	1..2457	/gene="phtB"		
	1..2457	/note="PhtB: protective pneumococcal antigen"		
	1..2457	/codon_start=1		
	1..2457	/translation="pneumococcal histidine triad protein B precursor"		
	1..2457	/protein_id="AAK06759.1"		
	1..2457	/db_xref="GI:12744741"		
	1..2457	/translation="MKINKYLAGSVAVLALSVGSYELGRYQAGQDKKESNRVAYIDG DOAGKAEPLTDEYSREGINAEQIVIKITDQGYVTSHGDHYHNGKYPYDAIIE ELLMKDPYQLKSDIUNELIKGVYKNGYVYVLDAAHADNRKKEIKQKQER SHNRADNAVAAAGRYTDDGYIENADIIEDTDGAYIVPHGDHYIIPKRL SASELAAEYWNKGKGRPSSSSNNANPAQRLSENHNLTVPTPHQNGENISL LRELAKPLSRVSDGLIFDPAQITSTRGAVAPHGNHYHIFPEQSELEKRIA RIILPYRNHWVPDRPEPSPQPSNPIDGKLKVEAKRVKGDGV PEENGVSRYIPAKDLSAETAGIDSKLAKQESLHKLTDLPSDRFYNKAYDL LARIHQDLNDKRGROVDFEALONLIERKQVSSDKVLVEDILAFIAPIRHPELGRP NAQITTDIDEIQVAKLAGKTAEDGYIFDPDITSDGDAVTPHMTSHWIKDSLS EAERAAQAAYAKEGLTTPSTQDSQNGTEAKGAELIYRVKAAKYPDLRMPNLOY TVEVKNGLIIPHYDHYNIKEWFDEGLIYAPQTEGLIEDLLATVKYVVEHPNRPHS DNGFGNASDHVQRNKQADNTQKTEPEERQTEKPEETREKPOSEKPSFKPT		

QY 160 GlnGlyArgThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
 Db 532 CAAGGACGTTATACACCGATGATGGTATATCTTCAATCATCTCATATCATGGAGC 591
 QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProIysAsnGlu 199
 Db 592 ACGGTGATGCTTATATCGTCTCAGCGGACCATTTACCAITACATTCCTAAGAAATGAG 651
 QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
 Db 652 TTATCAGCTAGCGAGTACTGCTCGAAGGCTATGGAATGGG----- 696
 QY 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239
 Db 697 -----AAGCAGGATCTCTGCTCTTCAAGTCTTAGTTATATGCA 738
 QY 240 SerValSerAspProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
 Db 739 AATCAGCTCAACCAAGATTGTGAGAGAACCAATCTGACTGTCACTCAACTTATCAT 798
 QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
 Db 799 CAA--AATCAAGGGGAAACATTTCAAGCCTTTACGTGAATGTATGCTAAACCCCTTA 855
 QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
 Db 856 TCAAGACGCCATGTGAATCTGATGGCTTATTTTCGACCCAGCGCAATCACAGTCCA 915
 QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
 Db 916 ACCGCCAGAGTGAGCTGCTCCCTCATGGTAACCATTACCACCTTTATCCCTTATGAACA 975
 QY 320 MetSerGluLeuGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
 Db 976 ATGCTCAATTGGAAAAACGAATTCCTGTATTAATCCCCCTTCGTATGCTTCAAAACAT 1035
 QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
 Db 1036 TGGGTACCAAGATTCAAGACCAAGAACCAAGTCCACACCGACTCCAGAACCTTAGTCCA 1095
 QY 360 GlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeu 379
 Db 1096 AGTCCGCAACAGCTCCAGCAATCCAAATTGAT-----GGGAAATTTGTCAAAGAGCT 1149
 QY 380 ValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPhe 399
 Db 1150 GTTCGAAAAGTAGCGATGCTTATGCTTTGAGAGAGATGAGTTCTCTGTTATATCCCA 1209
 QY 400 AlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGlu 419
 Db 1210 GCCAAGGATCTTTCAGCAGAAACAGCAGCAGGCAATGTATGCAAACTGGCCAGCAGGAA 1269
 QY 420 SerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPhe 439
 Db 1270 AGTTATCTCATAGTAGGAACCTAGAAACCTGACCTCCCACTAGTGTATGATCAGAAATTT 1329
 QY 440 TyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGly 459
 Db 1330 TACAATAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACTTGATATAAAGT 1389
 QY 460 ArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThr 479
 Db 1390 CCACAAGTTGATTTTGGAGCTTGGATAACCTTGTGGAACGACTCAAGGATGTCCTCACT 1449
 QY 480 AsnLysGluLysLeuValAspLeuLeuAlaPheLeuAlaProIleThrHisProGlu 499
 Db 1450 GATAAGTCAAGTTAGTGAAGATATCTTGTGCTTCTTAGCTCCGATTCGTATCCAGAA 1509
 QY 500 ArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGln 519
 Db 1510 CGTTAGGAAAAACCAATCGCAATTACCTACATGATGAGATGAGATCAAGTAGCCCAAG 1569
 QY 520 LeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSer 539

Db 1570 TTGGCAGGCAAGTACACAGCAGAGACGGTATATCTTTGATCCTCGTGATATAACCACT 1629
 QY 540 AspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAsp 559
 Db 1630 GATGAGGGGATGCTTATGTAATCTCCACATATGACCATAGCCACTGGAATTAAGAAAT 1689
 QY 560 SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeu 579
 Db 1690 AGTTTGTCTGAGCTGAGAGCGCGCAGCCAGGCTTATGCTTAAGAGAAAGTTTGACC 1749
 QY 580 ProProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIle 599
 Db 1750 CTTCTTCGACAGCATCAGGATTCAGAGAAATCTGAGGCAAAAGGACGACGAGTATC 1809
 QY 600 TyrAsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGlu 619
 Db 1810 TACAACCGGTGAAGACGCTAAGAAAGTGCACCTTGATCGTATGCTTACCAATCTCAA 1869
 QY 620 HisThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsn 639
 Db 1870 TATACCTGTAGAAGTCAAAACCGTAGTTTATCATCTCATCTATGACCATTTACCAATCTCAA 1929
 QY 640 IleLysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGlu 659
 Db 1930 ATCAAAATTTGAGTGGTTTGAAGAGGCTTTATGAGGCACCTAAGGGGTATACCTTTGAG 1989
 QY 660 AspLeuPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGlu 679
 Db 1990 GATCTTTGGGCTGTCAAGTACTATGTGGAACATCCAAACGAAACGTCGCAATTCAGAT 2049
 QY 680 AspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspPro 699
 Db 2050 AATGGTTTGTGTAACGCTAGCGACCTGTTCAAGAAACAAATAAGTGAAGCTGATACC 2109
 QY 700 AsnLysAsnPhenylAlaAspGluGlu-----ProValGluGluThr 713
 Db 2110 AATCAAAACGGAAAAACCAAGCGAGGAAACCTCAGACAGAAAAACCTGAGGAGAAACC 2169
 QY 713 ----- 713
 Db 2170 CCTCGAGAGAGAAACCGCAAGCGAGAAACCAAGAGTCTCCAAACCAACAGAGGAACCA 2229
 QY 714 -----ProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGln 730
 Db 2230 GAAGAATCACAGAGAAATCAAGAGAACCTCAGTGCAGACTGAAAGGTTGAAGAAAA 2289
 QY 731 LeuLysGluAlaGluValLeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAla 750
 Db 2290 CTGAGAGGCTGAAGATTTACTTTGAAAAATCCAGGATCCCAATTTCAAGTCCAATGCC 2349
 QY 751 ThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSer 770
 Db 2350 AAAGAGACTCTCAGAGATTAATAATAATTTACTATTGTCACCCAGCAGCAACAATACT 2409
 QY 771 IleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer 785
 Db 2410 ATTATGGCAGAGCTGAAAAACTATTGGCTTTTATTAAGAGAGAGT 2454

RESULT 23

BD268049

LOCUS

DEFINITION

Streptococcus pneumoniae protein and immunogenic fragments for

vaccines.

ACCESSION

BD268049

VERSION

BD268049.1 GI:33077817

KEYWORDS

JP 2002532561-A/7.

SOURCE

Streptococcus pneumoniae

ORGANISM

Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

Streptococcus

REFERENCE

1 (bases 1 to 2531)

AUTHORS

Johnson, L.S., Koenig, S. and Adamou, J.E.

TITLE Streptococcus pneumoniae protein and immunogenic fragments for
 JOURNAL Patient: JP 2002532561-A 7 02-OCT-2002;
 COMMENT MEDIMUNE INC
 OS Streptococcus pneumoniae
 PN JP 2002532561-A/7
 PD 02-OCT-2002
 PF 21-DEC-1999 JP 2000589215
 PR 21-DEC-1998 US 60/113048
 PI LESLIE S JOHNSON, SCOTT KOENIG, JOHN E ADAMOU
 PC A61K39/09, A61K38/00, A61P31/10, A61K37/02
 CC Streptococcus pneumoniae protein and immunogenic fragments for

CC vaccines
 FH Key Location/Qualifiers
 FT 1. .2531
 FT Location/Qualifiers
 FT /organism='Streptococcus pneumoniae'
 FT /organism='Streptococcus pneumoniae'
 FT /mol_type='genomic DNA'
 FT /db_xref='taxon:1313'

FEATURES
 source
 1. .2531
 /organism='Streptococcus pneumoniae'
 /mol_type='genomic DNA'
 /db_xref='taxon:1313'

ORIGIN
 Alignment Scores:
 Pred. No.: 6,99e-151 Length: 2531
 Score: 2786.50 Matches: 544
 Percent Similarity: 77.06% Conservative: 84
 Best Local Similarity: 66.75% Mismatches: 140
 Query Match: 66.90% Indels: 47
 DB: 6 Gaps: 7

US-09-765-271-56 (1-796) x BD268049 (1-2531)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnAsnArgValSer 19
 Db 61 TCCTATGAGCTTGGAGCTTACCAAGCTGGTCAGGATAGCAAGAGCTCTAATCAGGTGCT 120
 QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
 Db 121 TATATAGATGGTGATCAGGCTGGTCAAAAGGAGGAGAACTTGACACCATGAAGTCAGT 180
 QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
 Db 181 AAGAGGAGGGGATCAACGCCGCAAAATTTATTCAGATTAACGATCAAGGTTATG 240
 QY 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79
 Db 241 ACCTCTCATGGAGACCATATATCTATCTATATGCAAGGTTCTCTTATGATGCAATC 300
 QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
 Db 301 AGTGAAGAGCTCTCATGAAGATCGCAATATATCAGTTGAAGGATTCAGACATTCAT 360
 QY 100 GluValLysGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
 Db 361 GAAATCAAGGTTGGTATGTCATTAAGTAAACGTTAAATCTATGTTTACCTTAAGGAT 420
 QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
 Db 421 GCGGCTCATGGGATATATTCGCAAAAGAGAGATTAAACGTCAGAGAGGAGACGC 480
 QY 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaAtgSer 159
 Db 481 AGTCATAATCAATACTCA-----AGAGCAGATATGCTGTGTGTCAGCCAGAGCC 531
 QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
 Db 532 CAAGGAGCTTATACACGGATGATGGGTATATCTTCAATGCATCTGATATCATTTAGGAC 591
 QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
 Db 592 ACGGGTGATGCTTATATCGTTCTCTCAGCGGACCATACCATTCATTCCTAAGATGAG 651

QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
 Db 652 TTATCAGCTAGCGAGTTAGCTGCTCAGAGCAATTTGGAAATGGG-----696
 QY 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrPValPro 239
 Db 697 -----AAGCAGGGATCTCGTCTCTTCAAGTTCTAGTTATATGCA 738
 QY 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
 Db 739 AATCCAGCTCAACCAAGATGTCAGAGAACCAACATCTGACTGTCTCACTCACTCACTATCAT 798
 QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
 Db 799 CAA---AATCAAGGGGAAACATTTCAAGCTTTTACGTGAATTTGTATGCTAAACCTTAA 855
 QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
 Db 856 TCAGAACGCCATGCGAATCTGATGGCTTATTTTCGACCCAGCGCAATACCAAGTCGA 915
 QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
 Db 916 ACCGCCAGAGGTAGCTGCTCATGTTAACCATTACCATTTATCCCTTATGAAACA 975
 QY 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
 Db 976 ATGTCGAATTCGAAACCAAGATTCGTCGTATTTATTCCTCTCGTTATTCGTTCAACCAT 1035
 QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
 Db 1036 TGGGTACCAGATTCAAGACCAAGAACCAAGTCCACACCGACTCCAGAACCTAGTCCA 1095
 QY 360 GlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeu 379
 Db 1096 AGTCGCAACAGCTCCCAAGCAATCCAATTGAT-----GGGAAATTTGGTCAAAAGCT 1149
 QY 380 ValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPhe 399
 Db 1150 GTTCAAAAGTAGGCGATGTTATGCTTTGAGGAGAAATGGAGTTCTCTGTTATATCCCA 1209
 QY 400 AlaLysAspLeuProSerGluThrValLysAsnLeuLysSerLysLeuSerLysGlnGlu 419
 Db 1210 GCCAAGATCTTTCAGCAAAACAGCAGCAGCAGCTGATGATGATGATGATGATGATGAT 1269
 QY 420 SerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPhe 439
 Db 1270 AGTTTATCTCATAGCTAGGAACTAAGAAACTGACCTCCCATCTAGTGTATGATGATGAT 1329
 QY 440 TyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGly 459
 Db 1330 TACAATAAGGCTTATGACTTACTAGCAAGATTCACCAAGATTTACTTGTATATAAAGGT 1389
 QY 460 ArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThr 479
 Db 1390 CGCAAGTTGATTTTGGAGCTTTGGATAACCTGTTGCAAGCTCAAGGATGCTCAAGT 1449
 QY 480 AsnLysGluLysLeuValAspLeuLeuAlaPheLeuAlaProIleThrHisProGlu 499
 Db 1450 GATAAAGTCAAGTTAGTGAAGATATTTCTGCTCTTCTAGTCCGATTCCTCAGCA 1509
 QY 500 ArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGln 519
 Db 1510 CGTTTGAAGAAACCAATGCGCAATTTACCTACATGATGATGATGATGATGATGATGAT 1569
 QY 520 LeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSer 539
 Db 1570 TTGGCAGGCAAGTACACAGCAGAGAGCGTTATATCTTTGATCTCTGCTGATATAACCAT 1629
 QY 540 AspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAsp 559
 Db 1630 GATGAGGGGATGCTTATGTAACCTCCACATATGACCATGACCATGACCATGACCATGAC 1689
 QY 560 SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeu 579

```

Db 1690 AGTTTCTGAAGCTGAGAGCGGCGCCAGGCTTATCTRAAGAGAAAGTTTGACC 1749
Qy 580 ProProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIle 599
Db 1750 CCTCTCTCACAGACCATCAGGATTCAGGAATCTAGAGCAAAAGGAGCAGAGTATC 1809
Qy 600 TyrAsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGlu 619
Db 1810 TACAACCGMGTTGAAGCAGCTAAGAGGTTGCCATTTGATCGTATGCTTACATCTTCAA 1869
Qy 620 HisThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsn 639
Db 1870 TATCTGTAGAAGTCAAAACCGTATGTTAATCATACCTCATTTATGACCATACATAAC 1929
Qy 640 IleLysPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsn 659
Db 1930 ATCAAAATTTGAGTGTTCACGAGCCCTTTATGAGGACCTTAAAGGGGTATCTCTTGAG 1989
Qy 660 AspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsn 679
Db 1990 GATCTTTTGGGACGTCTCAAGTACTATGTGCAACATCCAAACGAAAGTCCGATTCAGAT 2049
Qy 680 AspGlyTyrGlyAsnAlaSerCluHisValLeuGlyLysLysAspHisSerGluAspPro 699
Db 2050 AATGTTTTTGTAGCGCTAGCGACCATGTTCAAGAAACAAATAATGTTCAAGCTCATACC 2109
Qy 700 AsnLysAsnPheLysAlaAspGluGlu-----ProValGluGluThr 713
Db 2110 AATCAAAACGGAACCAACGAGGAGGAGAACTCTCAGACAGAAACCTTGAGGAAGAAC 2169
Qy 713 ----- 713
Db 2170 CCTCGAGAAGAAACCGCAAGCGAGAAACCCAGAGTCTCCAAACCAACAGAGAACCA 2229
Qy 714 -----ProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGln 730
Db 2230 GAAGAATCACAGAGGAATCAGAAAGAACCTCAGGTCGAGACTGAAAGGTTGAAGAAAAA 2289
Qy 731 LeuLysGluAlaGluValLeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAla 750
Db 2290 CTGAGAGAGCTGAGAGTTTACTTGGAAATAATCCAGGATCAATATCAAGTCCAAATGCC 2349
Qy 751 ThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSer 770
Db 2350 AAAGAGAGCTCTCAGAGGATTAATAATAATTTACTATTTCGCACCCAGGACCAACATACT 2409
Qy 771 IleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer 785
Db 2410 ATTATGCGAGAGCTGAAAACTATTGGCTTTTATTAAAGGAGAGT 2454

RESULT 24
AR344448
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 2531)
AUTHORS
TITLES
JOURNAL
FEATURES
source
1. .2531
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:

```

```

Pred. No.: 6,99e-151 Length: 2531
Score: 2786.50 Matches: 544
Percent Similarity: 77.06% Conservative: 84
Best Local Similarity: 66.75% Mismatches: 140
Query Match: 66.90% Indels: 47
DB: 6 Gaps: 7

US-09-765-271-56 (1-796) x AR344448 (1-2531)
Qy 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnAsnArgValSer 19
Db 61 TCTATAGCTTGGAGCTTACCAAGCTGTCAGGATAAGAAAGAGTCTAATCGAGTTGCT 120
Qy 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
Db 121 TATATAGATGCTGATCAGGCTGTCAAAAGGCGAGAAACTTCACACCATGAGTCACT 180
Qy 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
Db 181 AAGAGGGAGGGGATCAACGCCGAAACAAATTTGTTATCAAGATTACGATCAAGTTATGTG 240
Qy 60 ThrSerHisGlyAspHisTyrHisTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyr 79
Db 241 ACCCTCTCATGAGACCATATATCTATCTATATATGTCAGAGGTTCTCTATGATGCCATC 300
Qy 80 SerGluGluLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
Db 301 AGTGAAGAGCTCTCATGAAAGATCCGAATTTATCAGTTGAAGGATTCAGACATTGCTCA 360
Qy 100 GluValLysGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysAsp 119
Db 361 GAAATCAAGGGTGTATGTCTATTAAGGTAAACGTAATAACTATGTTTACCTTAAGGAT 420
Qy 120 AlaAlaHisAlaAspAsnValAtqThrLysGluGlnIleAsnArgGlnLysGlnGluHis 139
Db 421 GCRCTCATCGGATATATTCGCAAAAGAGAGATTAACCTCAGAGCAGAGAACGC 480
Qy 140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
Db 481 AGTCATAATCAATACTCA-----AGAGCAGATTAATGCTGTGTCGAGCCAGAGCC 531
Qy 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleLeuAsp 179
Db 532 CAAGAGAGTTTATCAACGATGATGGTATATCTTCAATCATCTGATATCATTTAGGAC 591
Qy 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
Db 592 ACGGTGATGCTTATATCTGTTCTTCACGGGACCATTTACCATTCATTCCTAAGAATGAG 651
Qy 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
Db 652 TTATCAGCTAGCGAGTTAGCTGTCGAGAGCCCTATTGGAATGGG----- 696
Qy 220 AsnSerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239
Db 697 -----AAGCAGGAGTCTCGCTCTTCTTCAAGTTCTAGTTATAATGCA 738
Qy 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
Db 739 AATCCAGCTCAACCAAGATTGTCAGAGAACCATCTGACTGTCTCATCTTATCAT 798
Qy 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
Db 799 CAA---AATCAAGGGGAAACATTTCAAGCCCTTTTACGTTGATTTGCTAACCCTTA 855
Qy 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
Db 856 TCAGAACGCCATGTGAATCTGATGCGCTTATTTTCGACCCAGCGCAATCAAGTCA 915
Qy 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
Db 916 ACCGCCAGAGGTGAGCTGCTCTCATGTAACCATTTACCTTTTATCCCTTATGAACAA 975

```

```

320 MetSerGluLeuGluArgIleAlaArgIleProLeuArgTyrArgSerAsnHis 339
339 ATGTCGTAATGGAAAAACGAATGCTCGTATTATCCCTTCGTTATCGTTCAACCAT 1035
340 TrpValProAspSerArgProGluGluProSerProGlnProThrProGluProSerPro 359
359 ATGTCGTAATGGAAAAACGAATGCTCGTATTATCCCTTCGTTATCGTTCAACCAT 1035
1036 TGGGTACAGATTCAAGACAGAGAACCAAGTCCACACCGACTCCAGAACCTAGTCCA 1095
360 GlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeu 379
379 ATGTCGTAATGGAAAAACGAATGCTCGTATTATCCCTTCGTTATCGTTCAACCAT 1035
1096 AGTCGGCAACAGCTCCAGCAATCCATTCAT-----GGAAATTTGGTCAAAAGACT 1149
380 ValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPhe 399
399 ATGTCGTAATGGAAAAACGAATGCTCGTATTATCCCTTCGTTATCGTTCAACCAT 1035
1150 GTTCGAAAGTAGCGATGTTATGCTCTTCGAGGAGATGAGTTTCGCTTATATCCCA 1209
400 AlalysAspLeuProSerGluThrValLysAsnLeuGluSerLysSerLysGlnGlu 419
419 ATGTCGTAATGGAAAAACGAATGCTCGTATTATCCCTTCGTTATCGTTCAACCAT 1035
1210 GCCAAGGATCTTCAGCAAGAACACGACGACGATGATAGCAACTGGCCAAAGCAGAA 1269
420 SerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPhe 439
439 ATGTCGTAATGGAAAAACGAATGCTCGTATTATCCCTTCGTTATCGTTCAACCAT 1035
1270 AGTTTATCTCATAGCTAGGAATTAAGAACTAGCACTCCCATCTAGTATGATCGAGATTT 1329
440 TyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGly 459
459 ATGTCGTAATGGAAAAACGAATGCTCGTATTATCCCTTCGTTATCGTTCAACCAT 1035
1330 TACAATAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTTACTTCATATTAAGGT 1389
460 ArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThr 479
479 ATGTCGTAATGGAAAAACGAATGCTCGTATTATCCCTTCGTTATCGTTCAACCAT 1035
1390 CCACAAAGTTGATTTTGGAGCTTTGGTAACCTGTTGGAACGACTCAAGGATGCTCAAGT 1449
480 AsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGlu 499
499 ATGTCGTAATGGAAAAACGAATGCTCGTATTATCCCTTCGTTATCGTTCAACCAT 1035
1450 GATAAGTCAAGTTAGTGAAGATATTTCTGCTTCTAGCTCCGATTCGTCATCCAGAA 1509
500 ArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGln 519
519 ATGTCGTAATGGAAAAACGAATGCTCGTATTATCCCTTCGTTATCGTTCAACCAT 1035
1510 CGTTTGGAGAAACCAATGCGCAATTTACTCCTACCTGATGATGATGATCAAGTAGCCAG 1569
520 LeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleLeuSer 539
539 ATGTCGTAATGGAAAAACGAATGCTCGTATTATCCCTTCGTTATCGTTCAACCAT 1035
1570 TTGGCAGGAGTACACAGCAGAGAGCGTTATATCTTTGATCCTCGTATATAACCACT 1629
540 AspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisThrIleGlyLysAsp 559
559 ATGTCGTAATGGAAAAACGAATGCTCGTATTATCCCTTCGTTATCGTTCAACCAT 1035
1630 GATCAGGGGATGCTATGTAATCCACATATGACCATGACCATGCGATGATTAAGAAAGAT 1689
560 SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeu 579
579 ATGTCGTAATGGAAAAACGAATGCTCGTATTATCCCTTCGTTATCGTTCAACCAT 1035
1690 AGTTTGTCTGAAGCTGAGAGCGCGCCAGCGCTTATGCTTAAGAGAAAGGTTTGACC 1749
580 ProProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIle 599
599 ATGTCGTAATGGAAAAACGAATGCTCGTATTATCCCTTCGTTATCGTTCAACCAT 1035
1750 CCTCTCTGACAGCATCAGGATTCAGGAATATCTAGGCAAAAGAGCAGAGCTATC 1809
600 TyrAsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGlu 619
619 ATGTCGTAATGGAAAAACGAATGCTCGTATTATCCCTTCGTTATCGTTCAACCAT 1035
1810 TACAACCGMTGAAAGCAGCAGTAAAGAGTGCGCACTTGTATGCTTCAATCTTCA 1869
620 HisThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsn 639
639 ATGTCGTAATGGAAAAACGAATGCTCGTATTATCCCTTCGTTATCGTTCAACCAT 1035
1870 TATACCTGTAAGTCAAAACGGTAGTTTAAATCATCTCTATGATGACCATACCAATAC 1929
640 IleLysPheAlaThrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGlu 659
659 ATGTCGTAATGGAAAAACGAATGCTCGTATTATCCCTTCGTTATCGTTCAACCAT 1035
1930 ATCAATTTGAGTTGTTGACGAGCGCTTTATGAGGCACCTAAGGGGTATCTCTTAC 1989
660 AspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsn 679
679 ATGTCGTAATGGAAAAACGAATGCTCGTATTATCCCTTCGTTATCGTTCAACCAT 1035
1990 GATCTTTTGGCGACTGTCAAGTACTATATCGAATCATCCAAACGACGACCGCCATTCAGT 2049
680 AspGlyTyrPheAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspPro 699
699 ATGTCGTAATGGAAAAACGAATGCTCGTATTATCCCTTCGTTATCGTTCAACCAT 1035

```

```

2050 AATGGTTTGGTAACTAGCTAGCAGCCATGTTCAAAGAAACAAAATGGTCAAGCTGATACC 2109
700 AsnLysAsnPheLysAlaAspGluGlu-----ProValGluGluThr 713
2110 AATCAAAACGGAACCAAGCGAGGAGAAACCTCAGACAGAAACCTGAGGAAGAAACC 2169
713 ----- 713
2170 CCTCGAGAGAGAAACCGCAAGCGAGAAACCCAGAGTCTCCAAACCAACACAGAGAAACCA 2229
714 -----ProAlaGluProGluValProGlnValThrGluLysValGluAlaGln 730
2230 GAAGAATCACCAGAGGAATCAGAAAGAACCTCAGTCCGAGACTGAAAGGTTGAGAAAAA 2289
731 LeuLysGluAlaGluValLeuAlaLysValThrAspSerSerLeuLysAlaAsnAla 750
2290 CTGAGAGAGGCTGAAGATTACTTTGGAAAAATCCAGGATCCCAATATCAAGTCCCAATGCC 2349
751 ThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnSer 770
2350 AAAGAGACTCTCAGAGGATTAATAAATAATTTACTATTGGCACCAGGACCAACAATACT 2409
771 IleMetAlaGluAlaGluLysLeuAlaLeuLysGlySer 785
2410 ATTATGGCAGAGCTGAAATACTATTGGCTTTATTAAAGGAGT 2454

```

```

RESULT 25
SPNEU1901 232807 bp DNA linear HTG 12-JUL-2001
LOCUS Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.
DEFINITION AL449923
ACCESSION AL449923
VERSION AL449923.1 GI:11545148
KEYWORDS HTG: HTGS PHASE2.
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Humbert, Y.,
Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
Francesco, M., Polissi, A., Buell, G., Peger, G., Garcia, E., Peitsch, M.
and Garcia-Bustos, J.F.
Annotated draft genomic sequence from a Streptococcus pneumoniae
type 19F clinical isolate
Microb. Drug Resist. 7 (2), 99-125 (2001)
2 (bases 1 to 232807)
21335329
11442348
JOURNAL PUBMED
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 232807: contig of 232807 bp in length.
* Location/Qualifiers
1. 232807
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/serotype="19F"
/db_xref="taxon:1313"
/clone="G54"

```

ORIGIN

Alignment Scores:

Pred. No.: 4,18e-148 Length: 232807
 Score: 2781.00 Matches: 543
 Percent Similarity: 76.89% Conservative: 89
 Best Local Similarity: 66.06% Mismatches: 136
 Query Match: 66.77% Indels: 54
 DB: 2 Gaps: 7

US-09-765-271-56 (1-796) x SPNEU1901 (1-232807)

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnAsnArgValSer 19
 Db TCCTATGAACCTTGGCGCTTACCAAGCTGGTCAGGATAAGAAAGAGTCTATCGAGTTGCT 200412
 Qy 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
 Db TATATAGATGGTGCATCGGCTGCTCAAAAGGCGAGAAAATTTGACACAGATGAAGTCAGT 200472
 Qy 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
 Db AAGAGGAGGGGATCAACGCCGAACAAATTTGATCAAGATTACGGATCAAGGTATGTG 200532
 Qy 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIle 79
 Db ACCTCTCATGGAGACCATATATCTACTATAATGGCAAGGTTCTTATGATGCCATCATC 200592
 Qy 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
 Db AGTGAAGAGCTCCTCATGAAGATTCGAATTAATCAGTTGAAGATTTCAGACATTTGCAAT 200652
 Qy 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
 Db GAATCAAGGGTGGTATGTCATTAAAGTAACGGTAAATACTATGTTTACCTTAAGGAT 200712
 Qy 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
 Db GCAGCTCATGCGGATAATATTCGGCAAAAGAGAGATTAAACGTCAGAGCAGGAGCGC 200772
 Qy 140 SerGlnHisArgGluGlyThrProArgAspAspGlyAlaValAlaLeuAlaArgSer 159
 Db AGTCATAATCATAACTCA-----AGAGCAGATAATGCTGTGTCAGCGCAGAGCC 200823
 Qy 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
 Db CAAGGACGTACACACAGATGATGTTATCTTTAATGCTTCTGATATCATTTGAAGAT 200883
 Qy 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
 Db ACTGGTATGCTTATATCGTTCCTCATGGCAATCACTTCCACTATATTCCTAAGAGCTGAT 200943
 Qy 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
 Db TTATCTGCCAGTGAATAGTCTGCTGCCCAAGCCTATTGGAAATGGG----- 200988
 Qy 220 AsnSerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239
 Db -----AAGCAGGGATCTCGCTCTTCTCAAGTTCTAGTCATAATGCA 201030
 Qy 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
 Db AATCCAGCTCAACCAAGATTGTCAGAAACCAACAACTGCTGCTCACTCAACTTATCAT 201090
 Qy 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
 Db CAA---AATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTTGATGCTAAACCTTTA 201147
 Qy 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
 Db TCAGAACGCCATGTGAATCTGATGSCCTTATTTTCGACCCAGCGCAATTCAGATCGCA 201207
 Qy 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319

Db 201208 ACCCCAGAGGTAGCTGCTCCCTCATGGTAACCAATACCACCTTATCCCTTATGAACAA 201267
 Qy 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
 Db ATGTCCTGAATTTGAAACCAAGATTTGCTCGTATATTCCTTCCCTTCGTTATCGTTCAACCAT 201327
 Qy 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
 Db TGGGTACAGATTCAAGACCAGAACCAACCAAGTCCACACGCTCCAGAACCTAGTCCA 201387
 Qy 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373
 Db AGTCCGCAACTGCACCAATCTCCAACAGCTCCCAACCAATCCAAATGAT-----GAG 201441
 Qy 374 SerLeuValSerGlnLeuValArgLysValGlyGlyTyrValPheGluGluLysGly 393
 Db AATTTGGTCAAGAGAGCTGTTCCGAAAGTAGGCGATGTTATGCTTTGAGGAGAAATGGA 201501
 Qy 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413
 Db GTTTCCTGTTATATCCAGCCAAAGGATCTTTCAGCAGAAACAGCAGCAGCATTGATAGC 201561
 Qy 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433
 Db AACTGGCCCAAGCAGGAAGTTTATCTCATAGCTAGAGCTAAGAAACCTGACCTCCCA 201621
 Qy 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453
 Db TCTAGTGATCGAGATTTTACAATAAGGCTTATGACTTACTAGCAAGAAATCCCAAGAT 201681
 Qy 454 LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473
 Db TTACTTGATAATAAAGGTCGACAAAGTTGATTTTGGAGCTTTGGATAACCTGTTGGAACGA 201741
 Qy 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla 493
 Db CTCAGAGATGCTCAAGTGATTAAGTCAAGTTAGTGGAGATATCTTGGCTTCTTCTAGCT 201801
 Qy 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513
 Db CCGATTCTGTCATCCAGACGTTTAGCAAAACCAATTCGCAAAATTAACCTACACTGATGAT 201861
 Qy 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533
 Db GAGATTCAAGTAGCCCAAGTTGCGAGCAAGTACACAAACAGAACGCGTTATATCTTTGAT 201921
 Qy 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553
 Db CCTCGTGATATAACCACTGATGAGGGGATGCTATGTAACTCCACATATGACCCATAGC 201981
 Qy 554 HisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr 573
 Db CACTCGATTAAAAAAGATAGTTTGTCTGAAGCTGAAAGAGCGCAGCCCGAGGCTTATGCT 202041
 Qy 574 LysGluLysGlyIleLeuProSerProAspAlaAspValLysAlaAsnProThrGly 593
 Db AAGAGAAAGGTTTACCCCTCTTCGACAGACCATCAGGATTACGAGAAATACTAGGCA 202101
 Qy 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg 613
 Db AAGAGGAGCAAGACTATCTCAACCGCTGAAGCAAGCAGCTAAGAGAGTGCACCTGATCGC 202161
 Qy 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633
 Db ATGCCTTACAATCTCAATATACTGTAGAAGTCAAAACCGTAGTTTATATCATCTCAT 202221
 Qy 634 LysAspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAlaPro 653
 Db TATGACCATTAACCAATCAATCAAAATTTGAGTGGTTTGAAGAGGCTTTATGAGGACCT 202281
 Qy 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673

Db 202282 AAGGGTATACTCTTGAGGATCTTTGGCGACTGCTCAAGTACTATGTGGAACATCCAAAC 202341
Qy 674 GluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLys 693
202342 GAACGTCGGCATTGAGATAATGGTTTGGTAACGCTAGCGACCATGTTCAAGAACAACAAA 202401
Qy 694 AspHisSerGluAspProAsnLysAsnPhelLysAlaAspGlu----- 707
Db 202402 AATGGTCAAGCTGATACCAATCAAAACGGAACAAACCAACGAGAGAAACCTCAGACAGAA 202461
Qy 707 ----- 707
Db 202462 AAACCTGAGGAGAGAAACCCCTCGAGAGAGAGAAACCGAAACCGAGAAACAGAGTCTCCA 202521
Qy 708 -----GluProValGluGluThrProAlaGluProGluValProGlnValGlu 723
202522 AAACCAACAGACGAAACCAAGAGAGAAATCACCAGAGGAATCAGAAGAACCTCAGGTCGAG 202581
Qy 724 ThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAsp 743
202582 ACTGAAAGGTTGAAGAAACTGAGAGAGGCTGAAGATTCTTGGAAAAATCCAGGAT 202641
Qy 744 SerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeu 763
Db 202642 CCAATTATCAAGTCCCAATGCCAAAGAGACTCTCAGAGATTAAAAATAATTTACTATT 202701
Qy 764 GlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLys 783
Db 202702 GGCACCCAGGACAAACATATTATTATGGCAGAAAGCTGAAAAACTATTGGCTTTATTAAAG 202761
Qy 784 GlySer 785
Db 202762 GAGAGT 202767

Search completed: October 1, 2004, 10:19:24
Job time : 8737 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 1, 2004, 07:35:01 ; Search time 748 Seconds

(without alignments)
4520.810 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 4165

Sequence: 1 STELGLYQARTVKNRVS.....KLLALLKGNPSVSKEKIN 796

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying Chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool.h/US09765271/runat.30092004.113752.12270/app.query.fasta.1.967
-DB=N_Geneseq 29Jan04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS-human4.cdi
-LIST=90 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=25
-MODE=LOCAL -OUTPT=pt -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09765271 @CEN.1.1.513 @runat.30092004.113752.12270 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04.*

1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001as.*
5: Geneseq2001bs.*
6: Geneseq2002s.*
7: Geneseq2003as.*
8: Geneseq2003bs.*
9: Geneseq2003cs.*
10: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4163	100.0	2389	2	Aav27351 Streptoco
2	4163	100.0	2389	6	Abq84819 S. pneumo
3	4163	100.0	2389	9	Adc45136 S. pneumo
4	4163	100.0	2406	7	Abx06886 S. pneumo
5	4163	100.0	2451	3	Aaa47604 Recombina
6	4163	100.0	8195	2	Aav52227 Streptoco
7	4163	100.0	110000	7	Continuation (10 o
8	4163	100.0	110000	7	Continuation (11 o

9	3218	77.3	2523	3	AAA65731	Aaa65731 Streptoco
10	3218	77.3	2647	3	AAA65736	Aaa65736 Streptoco
11	3218	77.3	2647	6	ABK15103	Abk15103 DNA encod
12	2812	67.5	2478	3	AAA08557	Aaa08557 S. pneumo
13	2797.5	67.2	2457	7	ABX06885	Abx06885 S. pneumo
14	2795	67.1	2481	3	AAA05417	Aaa05417 Streptoco
15	2786.5	66.9	2531	3	AAA47605	Aaa47605 Recombina
16	2772	66.6	2517	7	ABX06705	Abx06705 S. pneumo
17	2772	66.6	2531	3	AAA47602	Aaa47602 Recombina
18	2772	66.6	110000	7	ABSS56454_08	Continuation (9 of
19	2750.5	66.0	2639	3	AAA65737	Aaa65737 Streptoco
20	2750.5	66.0	2639	6	ABK15104	Abk15104 DNA encod
21	2649.5	63.6	2290	6	AAV27356	Aav27356 Streptoco
22	2649.5	63.6	2290	6	ABQ84824	Abq84824 S. pneumo
23	2649.5	63.6	2290	9	ADC45146	Adc45146 S. pneumo
24	2431.5	58.4	2163	2	AAV25394	Aav25394 Streptoco
25	1774	42.6	3359	2	AAV52376	Aav52376 Streptoco
26	1591.5	38.2	3171	3	AAA65739	Aaa65739 Streptoco
27	1247	29.9	3120	3	AAA65730	Aaa65730 Streptoco
28	1247	29.9	5048	3	AAA65735	Aaa65735 Streptoco
29	1247	29.9	5048	6	ABK15101	Abk15101 DNA encod
30	1246	29.9	3117	7	ABX06706	Abx06706 S. pneumo
31	1222.5	29.4	6867	2	AAV52325	Aav52325 Streptoco
32	1216.5	29.2	110000	6	ABN71527_18	Continuation (19 o
33	1203.5	28.9	1455	3	AAA05473	Aaa05473 Streptoco
34	1203.5	28.9	1455	3	AAZ91804	Aaz91804 Streptoco
35	1203.5	28.9	1455	3	AAA65733	Aaa65733 Streptoco
36	1203.5	28.9	1455	3	AAA47603	Aaa47603 Recombina
37	1187.5	28.5	1342	2	AAV27414	Aav27414 Streptoco
38	1187.5	28.5	1342	6	ABQ84882	Abq84882 S. pneumo
39	1187.5	28.5	1342	9	ADC45262	Adc45262 S. pneumo
40	988	23.7	973	2	AAV52488	Aav52488 Streptoco
41	966.5	23.2	2528	3	AAA65738	Aaa65738 Streptoco
42	966.5	23.2	2528	6	ABK15105	Abk15105 DNA encod
43	950	22.8	2475	6	ABN66838	Abn66838 Streptoco
44	950	22.8	2478	5	AAAS00036	Aas00036 Streptoco
45	946	22.7	5215	3	AAA65740	Aaa65740 Streptoco
46	946	22.7	5215	2	AAK91105	Aak91105 Group B S
47	941	22.6	2472	3	AAA65741	Aaa65741 Streptoco
48	938	22.5	1398	6	ABN66839	Abn66839 Streptoco
49	937	22.5	2466	6	ABN69535	Abn69535 Streptoco
50	937	22.5	2466	6	ABN70334	Abn70334 Streptoco
51	937	22.5	110000	6	ABN71527_12	Continuation (13 o
52	934	22.4	2469	5	AAAS00038	Aas00038 Streptoco
53	929	22.3	2469	3	AAA05811	Aaa05811 Group B S
54	841	18.2	1146	3	AAA05814	Aaa05814 Group B S
55	770.5	18.5	1684	2	AAV52391	Aav52391 Streptoco
56	650	15.6	2721	8	ACF42880	Acf42880 S. pneumo
57	644	15.5	816	8	ACF42878	Acf42878 S. pneumo
58	644	15.5	816	8	ACF42879	Acf42879 S. pneumo
59	643	15.4	819	6	ABK15106	Abk15106 DNA encod
60	643	15.4	819	8	ACF42881	Acf42881 S. pneumo
61	628	15.1	504	3	AAA08556	Aaa08556 S. pneumo
62	546	13.1	492	2	AAV25393	Aav25393 Streptoco
63	512.5	12.3	1455	3	AAA05838	Aaa05838 Group B S
64	373.5	9.0	1137	6	ABN66837	Abn66837 Streptoco
65	304.5	7.3	841	2	AAV28529	Aav28529 S. pneumo
66	304.5	7.3	841	4	ABA76858	AbA76858 Streptoco
67	302	7.3	2631	6	ABN68134	Abn68134 Streptoco
68	302	7.3	110000	6	ABN71527_08	Continuation (9 of
69	261	6.3	1729	2	AAV42952	Aav42952 Streptoco
70	223	5.4	2376	6	ABN68135	Abn68135 Streptoco
71	223	5.4	2379	5	AAAS00037	Aas00037 Streptoco
72	221	5.3	567	7	ABX06707	Abx06707 S. pneumo
73	176.5	4.2	10242	4	ABL20787	AbL20787 Drosophil
74	176.5	4.2	20978	4	ABL20786	AbL20786 Drosophil
75	172	4.1	7446	7	ACA19708	AcA19708 Prokaryot
76	172	4.1	7446	9	ACF05848	Acf05848 Methicill
77	171	4.1	2685	7	ABT14858	Abt14858 Pathogen
78	170	4.1	7434	4	AAS52179	Aas52179 Staphyloc
79	170	4.1	7434	7	ABT14965	Abt14965 Pathogen
80	170	4.1	7434	7	ACF73459	Acf73459 Staphyloc
81	170	4.1	7437	4	AAS55232	Aas55232 Staphyloc

82 169 4.1 2685 7 ACF72781 Staphyloc
83 169 4.1 2730 7 ABZ22901 Staphyloc
84 169 4.1 4703 7 ADA89766 Staphyloc
85 168.5 4.0 3132 6 ABK48054 CDNA enco
86 168.5 4.0 5388 4 AB125979 Drosophil
87 166 4.0 3067 4 AAH54788 S. epider
88 166 4.0 15071 4 AAB53785 S. epider
89 164 3.9 2676 7 ACA19807 Prokaryot
90 163.5 3.9 4505 4 ABL17779 Drosophil

ALIGNMENTS

RESULT 1

AAV27351
ID AAV27351 standard; DNA; 2389 BP.

XX AC AAV27351;

XX DT 02-OCT-1998 (first entry)

XX DE Streptococcus pneumoniae SP0036 nucleotide.

XX KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX detection; pneumonia; otitis media; meningitis; ss.

XX OS Streptococcus pneumoniae.

XX FH Key Location/Qualifiers
XX CDS 2..2389

XX FT /*tag= a

XX FT /product= "SP0036"

XX FT /transl_except= {pos.1367..1369,aa:Xaa}

XX FT /note= "no stop codon given; Xaa is unspecified"

XX PN W09818930-A2.

XX PD 07-MAY-1998.

XX PF 30-OCT-1997; 97WO-US019422.

XX PR 31-OCT-1996; 96US-0029960P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Kunsch CA, Choi GH, Johnson LS, Hromockyj A;

XX WPI; 1998-272224/24.

XX DR P-ESDB; AAW55090.

XX PT Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
XX - or their epitope-containing fragments, useful in protective or
XX therapeutic vaccines, and for diagnosis.

XX PS Claim 1; Page 59; 118pp; English.

XX The present sequence encodes a protein from Streptococcus pneumoniae. The
XX nucleic acid sequence encoding the Streptococcus pneumoniae protein can
XX be useful in vaccines for inducing protective antibodies against
XX Streptococcus pneumoniae, for treatment or prevention of infection e.g.
XX pneumonia, otitis media or meningitis. Probes based on the nucleic acid
XX are used to detect Streptococcus infection (by usual hybridisation or
XX amplification methods), also for isolating Streptococcus genes or their
XX allelic variants. The protein can be used similarly to detect specific
XX antibodies in standard immunoassays, especially for diagnosing or
XX monitoring infections. Antibodies which bind the protein are used to
XX detect corresponding antigens, to purify the protein and for passive
XX immunisation (optionally coupled to a toxin). Vaccines are administered,
XX e.g. by injection, orally or through the skin, typically at 0.01-1000
XX (especially 10-300) mu g/ml per dose

XX Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 0 U; 1 Other;

SQ

Alignment Scores:

Pred. No.: 2e-282 Length: 2389
Score: 4163.00 Matches: 796
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.95% Indels: 0
DB: 2 Gaps: 0

US-09-765-271-56 (1-796) x AAV27351 (1-2389)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
DB 2 TCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAATATATCGTGTTCCTAT 61
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
DB 62 ATAGATGGAAACCAAGCGACGCAAAACCGAGAAATTTGACTCTCTGATGAGGTAGCAAG 121
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
DB 122 CGTGAGGATCAATGCTGAGCAATCGTCATCAAGATAACAGACCAAGGCTATGTCACT 181
QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleSer 80
DB 182 TCACATGGCGACCACTATCATTTACAATGGTAAGGTTCTTATGACGCTATCATCACT 241
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
DB 242 GAAGAATTACTCATGAAGATCCAAACTATAAGCTAAAGATGAGGATTTGTTAATGAG 301
QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
DB 302 GTCAGGGTGGATATGTTATCAAGGTAGATGAAAATACTATGTTTACCTTTAAGGATGCT 361
QY 121 AlaHisAlaAspAsnValArgThrLysGluIleLeuAsnArgGlnLysGlnGluHisSer 140
DB 362 GCCACGCGGATAACGTCCTGTACAAAAGAGAAATCAATCGACAAAAACAAGAGCATAGT 421
QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
DB 422 CAACATCGTGAAGTGGAACTCCAAGAAACGATGCTGTGTGCTTGGCAGTCGCGAA 481
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
DB 482 GGACGCTATACTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGATACT 541
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
DB 542 GGTGATGCTTATATCGTTCTCTCATGGAGATCATTTACCATTTACATTCCTTAAAGATGAGTTA 601
QY 201 SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
DB 602 TCAGCTAGCGAGTTGGCTGTGTCGAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAT 661
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
DB 662 TCAAGAACCCTATCCCGCAAAAATAGCGATAAACAACCTTCAAGAACAACTGGGTACCTTCT 721
QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
DB 722 GTAAGCAATCAGAACTACAAATACAAACAGCAACACAGCAACACACTAAACAGTCAA 781
QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
DB 782 GCAAGTCAAGTAAATGACATGATGAGTCTCTTGAACAGCTCTACAAACTGCCCTTTCAGT 841
QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
DB 842 CAACGACATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCACAAATCACAAAGTCGAACA 901
QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
DB 902 GCTAGAGTGTTCAGTGGCCACAGGAGATCATTTACCACTTCATCCCTTCTCTCAAAATG 961

QY 321 SerGluLeuGluGluArgIleAlaArgIleProLeuArgTyrArgSerAsnHisTrp 340
 Db TCTGAATTTGGAGAACGAAATCGCTGTAATATCCCTTCGTTATCGTTCAACATGG 1021
 QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
 Db GTACCAGATTCAGAGCCAGAACCAAGTCCACAAACCGACTCCGGAACCTAGTCCAGC 1081
 QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
 Db CCGCAACTCGCACCAAAATCTTAAATATAGACTCAAAATCTCTTTGGTTAGTCAGCTGTA 1141
 QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
 Db CGAAAAGTTGGGAAGATATGATTCCAGAGAAAGGCACTCTCTGTTAATGCTTTTGGC 1201
 QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
 Db AAAGATTTTACCATCTGAACCTGTTAAAAATCTTGAAGCAAGTTATCAAAACAAGAGAGT 1261
 QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
 Db GTTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCTCGTGCACCAAGAAATTTAT 1321
 QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
 Db GATAAAGCATATATCTGTTAACTGAGGCTCATTAAGCCTTGTTGNAATTAAGGGTCGT 1381
 QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
 Db AATTCTGATTTCCAGGCTTAGACAAATATTAGAAGCTTGAATGATGAATCGACTAAT 1441
 QY 481 LysGluLysLeuValAspAspLeuAlaPheLeuAlaProIleThrHisProGluArg 500
 Db AAAGAAATTTGGTAGATGATTTAATGGCAATCTCTAGCACCAATTAACCCATCCAGACGA 1501
 QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
 Db CTTGGCAACCAAAATCTCAAAATGAGTATACTGAGACGAGTTCGTATGCTCAATTA 1561
 QY 521 AlaAspLysTyrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
 Db GCTGATAAGTATACACGTCAGATGTTTACATTTTGTATGAACATGATATAATCAGTAT 1621
 QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560
 Db GAAGGAGATGCATATGTAACGCTTCATATGGGCCATAGTCACTGGATTGGAAAGATAGC 1681
 QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
 Db CTTTCTGATTAAGAAAGTTGAGCTCAAGCTATATAAGAAAGGTTATCTTACCT 1741
 QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600
 Db CCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTAC 1801
 QY 601 AsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
 Db AATCGTGTGAAGGGGAAACCAATTTCCACTCGTTCATATATGTTGGTTAGCAT 1861
 QY 621 ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle 640
 Db ACAGTTGAGGTTAAACCGTAATTTGATTTATCTCTATGAAGATCATACCATATATT 1921
 QY 641 LysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660
 Db AAAATTTGCTGGTTGATGATGATACACATACAAAGCTCCAAATGGCTATACCTTGGAGAT 1981
 QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
 Db TTTGTTGCGCAGATTAAAGTACTAGTAGAACACCCCTGACGACGTCACCATTTCTAATGAT 2041

QY 681 GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
 Db GGATGGGGCAATGCGCAGTGCAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAAT 2101
 QY 701 LysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720
 Db AAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAACACCTGCTGAGCCAGAGTCCCT 2161
 QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
 Db CAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAAAGCAGAAAGTTTTCCTTGGGAA 2221
 QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
 Db GTAAAGGATTCCTAGTCTGAAAGCAATGCAACAGAACTCTAGCTGGTTTTCGAAATTAAT 2281
 QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780
 Db TTGACTCTTCAAAATATGGATAACAATAGTATCATGCGAGAGCAGAAATAATTTACTTGG 2341
 QY 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796
 Db TTGTTAAAGGAGTAATCCTTCTATCTGTAAGTAAGGAAAAATAAAC 2389
 RESULT 2
 ABQ84819
 ID ABQ84819 standard; DNA; 2389 BP.
 XX
 AC ABQ84819;
 DT 04-SEP-2002 (first entry)
 XX
 DE S. pneumoniae SP036 nucleotide sequence SEQ ID NO:55.
 XX
 KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
 KW antibacterial; Streptococcal infection; detection; gene; ds.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN US2002061545-A1.
 XX
 PD 23-MAY-2002.
 XX
 PF 22-JAN-2001; 2001US-00765272.
 XX
 PR 30-OCT-1997; 97US-00961083.
 XX
 PA (CHOI/) CHOI G H.
 PA (KUNS/) KUNSCH C A.
 PA (BARA/) BARASH S C.
 PA (DILL/) DILLON P J.
 PA (DOUG/) DOUGHERTY B.
 PA (FANN/) FANNON M R.
 PA (ROSE/) ROSEN C A.
 XX
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
 PI Rosen CA;
 XX
 DR WPI; 2002-479261/51.
 DR P-PSDB; ABP54584.
 XX
 PT New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
 PT and for preventing or attenuating disease caused by Streptococcus
 PT infection.
 XX
 PS Claim 1; Page 27; 70pp; English.
 XX
 CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
 CC pneumoniae antigens have antibacterial activity and can be used in
 CC vaccines. The S. pneumoniae antigens can also be used to prevent or
 CC attenuate a Streptococcal infection in an animal. The polynucleotides
 CC encoding the S. pneumoniae antigens can be used to detect Streptococcus

CC nucleic acids. AB084905 to AB085130 represent primers used in the cloning
 CC of S. pneumoniae ORFs (open reading frames) which are used in an example
 CC from the present invention

XX Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 0 U; 1 Other;

Alignment Scores:
 Pred. No.: 2e-282 Length: 2389
 Score: 4163.00 Matches: 796
 Percent Similarity: 100.00% Conservations: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.95% Indels: 0
 DB: 6 Gaps: 0

US-09-765-271-56 (1-796) x AB084819 (1-2389)

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVallysGluAsnAsnArgValSerTyr 20
 Db 2 TCTTACGAGTTGGGACTGCTCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTAT 61
 Qy 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
 Db 62 ATAGATGAAACACAGCGACGCAACAAACGGAGAAATTCACCTCCGTGAGGTAGCAAG 121
 Qy 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
 Db 122 CQTGAAGAAATCAATGCTGAGCAATCGTCATCAGATTAACAGCAAGGCTATGCTACT 181
 Qy 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleLeuSer 80
 Db 182 TCACATGCGCACCATCATATATTACAAATGTAAGTTCCTTATGAGCTATCATCATCAGT 241
 Qy 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
 Db 242 GAAGAAATTAATCATGAAAGATCCAAACATATAAGCTAAAGATGAGGATATTTGTAATGAG 301
 Qy 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrValLysAspAla 120
 Db 302 GTCAGGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGCT 361
 Qy 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnHisSer 140
 Db 362 GCCACGCGGATACGCTCGTACAAAAGAGGAAATCAATCGACAAACAAAGAGCATAGT 421
 Qy 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaIleAlaArgSerGln 160
 Db 422 CAACATCGTGAAGTGGAACTCCAGAAACGATGGTGGTGGCTTGGCAGCTTCGCA 481
 Qy 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
 Db 482 GGACGCTATCTACAGATGATGTTATATCTTTAATGCTTCGTATATCATAGAGGATCT 541
 Qy 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
 Db 542 GGTGATGCTTATATGCTTCCTCATGGAGATCAATACCATTCATTCCTAAGAATGAGTTA 601
 Qy 201 SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyValGlyAsnLeuSerAsn 220
 Db 602 TCAGCTAGCGAGTTGGCTGCTCGAGAGCTTCCTCATCTGCTGGTGGAGAAATCTGTCAT 561
 Qy 221 SerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
 Db 662 TCAAGAACCTATCGCCGCAAAATAGCGATAACACATTCAGAACAACTGGGTACCTTCT 721
 Qy 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
 Db 722 GTAAGCAATCCAGAACTACAATACTAACAAGCAACAAACAGCAACACTAACACTCAA 781
 Qy 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
 Db 782 GCAAGTCAAGTAATGACATTGATGCTCTTCAACACAGCTCTACAACTGCTCTTGGT 841
 Qy 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300

Db 842 CAACGACATGTAGAATCTGATGGCTTGTCTTTGATCCAGCACAAATCACAAAGTCGAACA 901
 Qy 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
 Db 902 GCTAGAGGTGTGAGTGGCCACACGAGATCATTACCACCTTCATCCCTTACTCTCAATG 961
 Qy 321 SerGluLeuGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340
 Db 962 TCTGAATTTGGAGAACGATCGCTCGTATTATTCCTCTGTTATTCGTTCAACACATTGG 1021
 Qy 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
 Db 1022 GTACCAGATTCAAGGCCAGAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGGC 1081
 Qy 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
 Db 1082 CGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGTA 1141
 Qy 381 ArgLysValGlyGlyTyrValPheGluGlyLysGlyIleSerArgTyrValPheAla 400
 Db 1142 CGAAAGTTGGGAGGATATGATTTCGAAGAAAGGCACTCTCTCGTTATGCTCTTGG 1201
 Qy 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
 Db 1202 AAAGATTTTACCATCTGAAACTGTTAAAAATCTTGAAGCAAGTATTATCAAAACAGAGAGT 1261
 Qy 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
 Db 1262 GTTTCACACACTTTAACTGCTTAAAGAAATGTTGCTCTCTGAGCAAGAAATTTAT 1321
 Qy 441 AspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
 Db 1322 GATTAAGCATATAATCTCTTAACCTGAGGCTCATTAAGCCTTGTGTTGNAATAAAGGTCGT 1381
 Qy 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
 Db 1382 AATCTGATTTCCAAAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAAT 1441
 Qy 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
 Db 1442 AAAGAAATTTGGTATAGATGATTTATTTGGCATTCCTAGCACCAATTTACCCATCCAGAGCA 1501
 Qy 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
 Db 1502 CTTGGCAAAACCAATTTCTCAATTCAGTATATCTGAAGACGAAGTTCGTATTCTCAATTA 1561
 Qy 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
 Db 1562 GCTGATAAGTATACAACTGATGCTTACATTTTGTGTAACATGATATAATCATGAT 1621
 Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560
 Db 1622 GAAGAGATGCATATGTAACGCTCATATGGCCATAGTCACTGGATGGAAAGATAGC 1681
 Qy 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
 Db 1682 CTTTCTGATAAGCAAAAGTTGCAGCTCAAGCCTATCTATAAAGAAAAAGGTATCTTACCT 1741
 Qy 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600
 Db 1742 CCATCTCAGACGCGAGATGTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTAC 1801
 Qy 601 AsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
 Db 1802 AATCGTGTGAAGGGGAAAAACGAATTCACCTCGTTCGACTTCCATATATGTTGAGCAT 1861
 Qy 621 ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle 640
 Db 1862 ACAGTGTAGGTAAAAACGGTAAATTTGATTTATTCCTCATAGGATCATTACCATAATAT 1921
 Qy 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660

Db 782 GCAAGTCAAGTAATGACATTTGATGCTCTTGAAACAGCTCTACAACTGCTTTGAGT 841
 QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
 Db 842 CAACGACATGTAGAATCTGATGCGCTTGCTTTGATCCAGCACAAATACAAAGTCGAACA 901
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
 Db 902 GCTAGAGGTGTTGTCAGTGCACACGAGATCATACCACTTCATCCCTTACTCTCAATG 961
 QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340
 Db 962 TCTGAATTCGAAGACGAATCGCTGATATATATTCCTTCCTTCGTTATCCAAACCAATG 1021
 QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
 Db 1022 GTACCAAGTTCAGGCCACCAACCAAGTCCACACCCAGCTCCGGAACCTAGTCAGGC 1081
 QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
 Db 1082 CCGCAACCTGCACCAATCTTAAATATAGACTCAAAATCTCTTCTGTTAGTCAGCTGTA 1141
 QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
 Db 1142 CGAAGATTCGGAGAGATATGATTCGAAGAAAGGGCATCTCTGTTATGCTTTGCG 1201
 QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
 Db 1202 AAAGATTTACCATCTGMAACTGTAAATCTTGAAGCAAGTATCAAAACCAAGAGT 1261
 QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
 Db 1262 GTTTCACACACTTTAACTGCTAAAAAGAAAAATGTTGCTCTCTGTCACCAAGAAATTTAT 1321
 QY 441 AspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
 Db 1322 GATAAAGCATATATCTGTTAACTGAGGCTCTAAAGCCTTGTGTTGNAATTAAGGTCGT 1381
 QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
 Db 1382 AATCTGATTTCCAAAGCTTAGACAAATTAATAGAACGCTTGATGAATGATCGACTAAT 1441
 QY 481 LysGluLysLeuValAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
 Db 1442 AAAGAAAAATTTGGTAGATGATTTATGTCATCTTAGCACCATAATACCCTCAGAGGGA 1501
 QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
 Db 1502 CTGGCAACCAATCTCAATTCAGTATCTAGAACGAAAGTTCTGTTATGCTCAATTA 1561
 QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
 Db 1562 GCTGATAAGTATACAACTCAGATGTTTACATTTTGTATGAACATGATATATATCAGTAT 1621
 QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560
 Db 1622 GAAGGAGATGATATGAACCCCTCATATGGGCCATAGTCTCAGTGGATGGAAGATAGC 1681
 QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
 Db 1682 CTTTCTGATAAGAAAGTTGAGCTCAAGCCTATATAAGAAAAAAGGATCTCTACCT 1741
 QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600
 Db 1742 CCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTATAC 1801
 QY 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
 Db 1802 AATCGTGTGAAGGGGAAAAACGAATTCATCTGTTGACCTCCATATATATGTTGAGCAT 1861
 QY 621 ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle 640
 Db 1862 ACAGTTGAGTTAAAAACGGTAATTTGATTTATCTCTATAAGGATCATTAACATAATAT 1921

QY 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660
 Db 1922 AAATTTGCTTGGTTTGTATGATCACATACAAAGCTCCAAATGGCTATACCTTTGGAAGAT 1981
 QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
 Db 1982 TTGTTTGGCAGCATTAAGTACTACGTAGAACACCTTGACGACGTCCACATTTCAATGAT 2041
 QY 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
 Db 2042 GGATGGGGCAATGCCAGTGAGCATGTGTAGGCAAGAAAGACACACAGTGAAGATCCAAAT 2101
 QY 701 LysAsnPheLysAlaAspGluProValGluGluThrProAlaGluProGluValPro 720
 Db 2102 AAGAATCTCAAGCGATGAGAGCCAGTAGAGGAACACCTGTGAGCCAGAGTCCCT 2161
 QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
 Db 2162 CAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGCAGAGTTCCTGCTGCGAAA 2221
 QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
 Db 2222 GTAACGGATTCAGTCTGAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAAATAT 2281
 QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780
 Db 2282 TTGACTCTCTCAATTTATGATTAACATAGTATCATGGCAGAGCAGAAAAATTTACTTGG 2341
 QY 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796
 Db 2342 TTGTTAAAGGAAGTAACTCTTCATCTGTAGTAAGGAAAAATAAAC 2389

RESULT 4

ABX06886
 ID ABX06886 standard; DNA; 2406 BP.
 AC ABX06886;
 DT 27-OCT-2003 (revised)
 DT 11-FEB-2003 (first entry)
 XX S. pneumoniae type 4 strain coding region #1174.
 DE Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
 KW ear infection; antiinflammatory; antibacterial; immunostimulant;
 KW auditory; respiratory; gene therapy; vaccine.
 XX Streptococcus pneumoniae; type 4 strain.
 XX WO200277021-A2.
 XX 03-OCT-2002.
 XX 27-MAR-2002; 2002WO-IB002163.
 XX 27-MAR-2001; 2001GB-00007658.
 XX (CHIR-) CHIRON SPA.
 XX (GENO-) INST GENOMIC RES.
 FI Masignani V, Tettelin H, Fraser C;
 XX WPI; 2003-040579/03.
 DR P-PSDB; ABU01598.
 XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
 XX ear infection.
 PS Claim 6; SEQ ID NO 2347; 56pp; English.

CC The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC AB556454. Also included are an antibody which binds one of the proteins,
 CC treating a patient by administering the protein, DNA or antibody (in a
 CC composition), a kit comprising first and second primers, which are the
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a
 CC sequence not defined in the specification, for amplifying a target
 CC sequence contained within a Streptococcus nucleic acid sequence, where
 CC the first primer is substantially complementary to the target sequence
 CC and the second primer is substantially complementary to the complement of
 CC the target sequence, and where the parts of the primers having
 CC substantial complementarity define the termini of the target sequence to
 CC be amplified, assay comprising contacting a test compound with the
 CC protein, and determining whether the test compound binds to the protein
 CC and a Streptococcus pneumoniae bacterium, where one or more genes
 CC encoding the proteins has been rendered inactive. The proteins, nucleic
 CC acid molecules, antibody and compositions are useful as medicaments for
 CC treating or preventing a disease or infection due to streptococcus
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
 CC media or ear infection. They are also useful in developing vaccines,
 CC diagnostics and antibiotics. The methods are useful for identifying
 CC immunodominant proteins. The present sequence is one of the 2489
 CC identified coding region from the genomic sequence. Note: The sequence
 CC data for this patent did not form part of the printed specification, but
 CC was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to
 CC standardise OS field)

XX Sequence 2406 BP; 834 A; 461 C; 490 G; 621 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,01e-282 Length: 2406
 Score: 4163.00 Matches: 795
 Percent Similarity: 99.87% Conservative: 0
 Best Local Similarity: 99.87% Mismatches: 1
 Query Match: 99.95% Indels: 0
 DB: 7 Gaps: 0

US-09-765-271-56 (1-796) x ABX06886 (1-2406)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
 DB 19 TCTTACGAGTGGGACTGTATCAAGCTAGACGGTTAAGGAAATAATCGTGTTCCTAT 78
 QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
 DB 79 ATAGATGGNAACAACGCGACGCAAAAACGGAGAAATTGACTCCTGATGAGGTAGCAAG 138
 QY 41 ArgGluGlyIleAsnAlaGluGlnIleValLysLysLysLysLysLysLysLysLys 60
 DB 139 CGTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGCAAGAGCTATGTCACT 198
 QY 61 SerHisGlyAspHisThrHisTyrTyrAsnGlyLysValProTyrAspAlaIleLysSer 80
 DB 199 TCACATGGCGACCACTATCATATTACAAATGTAAGGTTCCTTATGACGCTATCATCACT 258
 QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
 DB 259 GAAGAATTAATCATGAAAGATCCAACTATTAAGCTAAAGATGAGATATTGTTAATGAG 318
 QY 101 ValLysGlyGlyTyrValLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
 DB 319 GTCAAGGGTGGATATGTTATCAAGGTAGATGAAATACTATGTTTACCTTAAGGATGCT 378
 QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnHisSer 140
 DB 379 GCCACGCGGATACGTCGTCGTAACAAAGAGGAATCAATCGACAAACAAAGAGCATAGT 438
 QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyValAlaLeuAlaArgSerGln 160

DB 439 CAACATCGTAGGTTGGAACTCCAAAGAACGATGGTGGCTGTTGCCCTTGGCAGTTCGCAA 498
 QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
 DB 499 GGACGCTATACATACAGATGATGGTTATATCTTAAATGCTTCTGATATATATGAGATACT 558
 QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
 DB 559 GGTGATGCTTATATGTTCTTCATGGAGATCATTAACCAATTATCATTCCTTAAGNAATGAGTTA 618
 QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
 DB 619 TCAGCTAGCGAGTTGGCTGCTGCAGAGCCTTCTCTATCTGGTTCGAGGAAATCTGTCAAA 678
 QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
 DB 679 TCAAGAACCTATCGCCGCAAAATAGCGATAACACTTCAAGAACCAAACTGGGTACTCTTCT 738
 QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
 DB 739 GTAAGCAATCCAGGAACCTACAAATACTAACACAGCAACACAGCAACACTAACAGTCAA 798
 QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
 DB 799 GCAAGTCAAAAGTATGACATGATGATAGTCTCTTTGAAACAGCTCTACAAACCTGCTTGA 858
 QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
 DB 859 CAACGACATGTAGAATCTGTCTTTGATCCAGCACAAATCAACAGTCCGACACA 918
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
 DB 919 GCTAGAGGTGTTGCACTGCCACAGAGATCATTAACCACTTCATCCCTTACTCTCAATG 978
 QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340
 DB 979 TCTGAATTTGGAAGAACGAATCGTCTGTTATTTCCCTTCGTTATCGTTTCAAAACCATGG 1038
 QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
 DB 1039 GTACCAGATTCAAGGCCAGAACCAACCAAGTCCACAAACCGACTCCGGAACTAGTCCAGGC 1098
 QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
 DB 1099 CCGCAACCTCGCACCAAACTTTAAATAGACTCAATTTCTCTTTGGTTAGTCAGTGGTA 1158
 QY 381 ArgLysValGlyGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
 DB 1159 CGAAAAGTTGGGAAAGGATATGTTATTCGAAAGAAAGGGCATCTCTCGTTATGCTTTGGC 1218
 QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
 DB 1219 AAAAGATTTTACCATCTGAAACCTGTTAAATACTTTGAAAGCAAGTTATCAAAAACAGAGAT 1278
 QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
 DB 1279 GTTTTCAACACTTTTAACTGCTAAAAAGAAATGTTGCTCTCTCGTACCAAGATTTTAT 1338
 QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
 DB 1339 GATAAAGCATATAATCTGTTAATCTGAGGCTCATAAAGCCTTGTGTTGAAAATAAGGGTCGT 1398
 QY 461 AsnSerAspPheGluAlaLeuAspLysLeuGluArgLeuAsnAspGluSerThrAsn 480
 DB 1399 AATTCTGATTTTCCAGCCCTTAGACAAATTTATTAGAACCGCTTGAATGATGAATCGACTAAT 1458
 QY 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
 DB 1459 AAAGAAAATTTGGTAGATGATTTTATTTGGCATTTCTTAGACCAATTAACCATCCAGAGCGA 1518
 QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
 DB 1519 CTTGGCAACCAAAATTTCTCAATTTGAGTATATCTGAAGACGAAGTTCTGTTATGCTCAATTA 1578

QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleLeuSerAsp 540
 DB 1579 GCTGATAGTATACAAAGTCAAGTGGTTCATNTTTGATGACATGATATATCAATGATGAT 1638
 QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560
 DB 1639 GAAGGAGATGCATATGTAAGCCTCATATGGCCATAGTCACTGGATTGGAAGATAGC 1698
 QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
 DB 1699 CTTTCTCATAGGAAAGTTCAGCTCAAGCCTATATCAAGAAAGATGATCTACCT 1758
 QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600
 DB 1759 CCATCTCCAGCAGATGTTAAGCAATCCAACTGAGATAGTGGAGCTATTATAC 1818
 QY 601 AsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
 DB 1819 AATCGTGTAAAGGGGAAACGAATCCACTCGTTCATATATGTTGAGCAT 1878
 QY 621 ThrValGluValLysAsnGlyAsnLeuIleLeuProHisLysAspHisTyrHisAsnIle 640
 DB 1879 ACAGTTCAGGTTAAAGCGGTAATTCATATTCCTCATAGGATCATACCAATAATATT 1938
 QY 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660
 DB 1939 AAATTTGCTGTTGATGATCACATACAAAGTCAATGCTATATCTTGGAGAT 1998
 QY 661 LeuPheAlaThrLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
 DB 1999 TTGTTTCGACGATTAAGTACTAGTAGAACACCCCTGACGAACGTCACATCTCTAATGAT 2058
 QY 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysAspHisSerGluAspProAsn 700
 DB 2059 GGATGGGCAATGCCAGTGCAGTGTGTAGGCAAGAAAGACCACTGAGATCCCAAT 2118
 QY 701 LysAsnPheLysAlaAspGluProValGluThrProAlaGluProGluValPro 720
 DB 2119 AAGAACTTCAAGCGGATGAAGCCAGTAGAGGAACACCTGCTGAGCAGAGTCCCT 2178
 QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
 DB 2179 CAAGTAGAGACTGAAAGTAGAGCCCACTCAAGAGCAGAAAGTTTGTCTGCGAAA 2238
 QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
 DB 2239 GTAACGATTTCTAGTCTGAAAGCCAAATGCACAGAAACTCTAGCTGTTTACGAAATPAT 2298
 QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780
 DB 2299 TTGACTCTTCAATTTATGATGATAACAATAGTATCATGCGCAGACGAGAAATTTACTTGG 2358
 QY 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796
 DB 2359 TTGTTAAAGGAAGTATCTTCTATCTGTAAAGTAAAGGAAAAATAAAC 2406

RESULT 5

AAA47604

ID AAA47604 standard; DNA; 2451 BP.

XX AAA47604;

AC AAA47604;

XX 20-OCT-2000 (first entry)

XX Recombinant variant of Sp36 gene (Sp36A) of *S. pneumoniae*.

DE Streptococcus pneumoniae; infection; vaccine; coiled coil region;

XX histidine triad residue; Sp36; antibody; otitis media;

XX nasopharyngeal infection; bronchial infection; bronchitis; sepsis;

XX meningitis; lobar pneumonia; ds.

XX Streptococcus pneumoniae.

XX Key Location/Qualifiers
 FH 1..2451
 FT /*tag= a
 FT /product= "Sp36A polypeptide"
 XX WO200037105-A2.
 XX 29-JUN-2000.
 XX 21-DEC-1999; 99WO-US030390.
 XX 21-DEC-1998; 98US-0113048P.
 XX (MEDI-) MEDIMUNE INC.
 XX Johnson LS, Koenig S, Adamou JE;
 XX WPI; 2000-452129/39.
 XX P-PSDB; AAB01468.
 XX Vaccine useful for prophylaxis and treatment of pneumococcal infections
 PT such as otitis media, nasopharyngeal and bronchial infections, comprises
 PT Streptococcus pneumoniae proteins.
 XX Disclosure; Page 64-65; 70pp; English.
 XX Although a number of proteins have been suggested as being involved in
 CC the pathogenicity of Streptococcus pneumoniae, there still remains a need
 CC to identify polypeptides having epitopes in common from various strains
 CC of *S. pneumoniae* in order to utilize such polypeptides in vaccines to
 CC protect against a wide variety of *S. pneumoniae*. New vaccine compositions
 CC are described which comprise a Streptococcus pneumoniae polypeptide (or
 CC fragments) of 80 - 680 amino acids in length that comprise at least one
 CC histidine triad residue (HxxHxH) or a coiled-coil region, or an antibody
 CC directed against these features. The vaccine is useful in protecting
 CC against infection by Streptococcus pneumoniae. The vaccine composition
 CC comprising antibodies to is useful for passive immunization for treating
 CC Pneumococcal infections which includes otitis media, nasopharyngeal and
 CC bronchial infections
 XX SQ Sequence 2451 BP; 849 A; 467 C; 499 G; 635 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.: 2,06e-282 Length: 2451
 Score: 4163.00 Matches: 796
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.95% Indels: 0
 DB: 3 Gaps: 0

US-09-765-271-56 (1-796) x AAA47604 (1-2451)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
 DB 61 TCTTACGAGTTGGCACTGATCAAGCTAGAACGGTTAAGGAAAAATAATCGTGTTCCTAT 120
 QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
 DB 121 ATAGATGGAAAAACAAGCGACGCAAAAACGGAGAATTTGACTCTCTGATGAGGTAGCAAG 180
 QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
 DB 181 CGTGAAGGAATCAATGCTGAGCAATCGTCATCAAGATACACAGCCAGGCTATGTCAT 240
 QY 61 SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleLeuSer 80
 DB 241 TCACATGGGACCACTATCATATTACAAATGGTAAAGTTCCTTATACGCTATCATCAGT 300
 QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
 DB 301 GAAGAAATTACTCATGAAGATCCAACTATAAGCTAAAAGATGAGGATATTGTTAATGAG 360

23-OCT-1998 (first entry)
 Streptococcus pneumoniae genome fragment SEQ ID NO:94.
 Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 computer readable medium; vaccine; pharmaceutical composition; ds.
 Streptococcus pneumoniae.
 WO9818931-A2.
 07-MAY-1998.
 30-OCT-1997; 97WO-US019588.
 31-OCT-1996; 96US-0029960P.
 (HUMA-) HUMAN GENOME SCI INC.
 Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;
 Dougherty BA;
 WPI; 1998-272225/24.
 Computer-readable medium with recorded Streptococcus pneumoniae
 polynucleotide sequences - useful in diagnostic kits and assays, and
 pharmaceutical compositions and vaccines for Streptococcus pneumoniae.
 Claim 1; Page 727-732; 1409pp; English.
 The present invention describes a computer readable medium which has the
 nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded
 on it, or a representative fragment or a sequence at least 95% identical
 to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
 to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus
 pneumoniae. The present invention also describes an isolated nucleic acid
 molecule encoding a homologue of any of the fragments of the S. pneumoniae
 genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
 by a process comprising: (a) screening a genomic DNA library using as a
 probe a target sequence defined by any of the sequences in SEQ ID NO:1 to
 391, identifying members of the library which contain sequences that
 hybridize to the target sequence and isolating the nucleic acid molecules
 from the members; or (b) isolating mRNA, DNA or cDNA produced from an
 organism, amplifying nucleic acid molecules whose nucleotide sequence is
 homologous to amplification primers derived from the fragment of the S.
 pneumoniae genome to prime the amplification and isolating the amplified
 sequences. The computer readable medium can be used in a computer-based
 system for identifying fragments of the S. pneumoniae genome of
 commercial importance, or expression modulating fragments of the S.
 pneumoniae genome. Products from the present invention can be used in
 diagnosis kits and assays, and pharmaceutical compositions and vaccines
 for S. pneumoniae
 XX SQ Sequence 8195 BP; 2688 A; 1622 C; 1777 G; 2105 T; 0 U; 3 Other;
 Alignment Scores:
 Pred. No.: 8,76e-282 Length: 8195
 Score: 4163.00 Matches: 795
 Percent Similarity: 99.87% Conservative: 0
 Best Local Similarity: 99.87% Mismatches: 1
 Query Match: 99.95% Indels: 0
 DB: 2 Gaps: 0
 US-09-765-271-56 (1-796) x AAV52227 (1-8195)
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
 Db 3054 TCTTACGAGTTGGAGCTATGATCAAGCTAGAACGGTTAAAGGAAAATAATCGTGTTCCTAT 3113
 QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
 Db 3114 ATAGATGGGAAAACAAAGCGCAACAAACGAGAAATTTGACTCCTGATGAGGTAGCAAG 3173

QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
 Db 3174 CGTAGGAATCAATCGTGTAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGCTACT 3233
 QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleSer 80
 Db 3234 TCACATGCGCACCATCATCATTTACCAATGTTAGGTTCCCTTATGACGCTATCATCAT 3293
 QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
 Db 3294 GAAGAAATTACTCATGAAAGATCCAAACATATTAAGCTTAAAGATGAGATATGTTAATGAG 3353
 QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrValLysAspAla 120
 Db 3354 GTCAGGGTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACTTAAAGATGCT 3413
 QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnLysHisSer 140
 Db 3414 GCCCAGCGGATACGTCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAGT 3473
 QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
 Db 3474 CAACATCGTGAAGGTGGAACTCCAAAGAACGATGCTGTGCTTGGCAGCTTCGCAA 3533
 QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
 Db 3534 GGAAGCTATACTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGTATCT 3593
 QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
 Db 3594 GGTGATGCTTATATCGTCTCATGGAGATCAATTACCATTCATCTCTAAGAAATGAGTTA 3653
 QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyValAsnLeuSerAsn 220
 Db 3654 TCAGCTAGGAGTGGCTGCTGCAGAGACCTTCCTATCTGGTCGAGGAAATCTGTCAAT 3713
 QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
 Db 3714 TCAAGAACCTATCGCGCAAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTCT 3773
 QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
 Db 3774 GTAAGCAATCCAGGAACCTACAAATACTAACACAGCAACACAGCAACACTAACAGTCAA 3833
 QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
 Db 3834 GCAAGTCAAGTAATGACATTCATGATCTCTTTGAACACCTCTACAAACTGCTTTGAGT 3893
 QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
 Db 3894 CAACGACATGTAGAATCTGATGGCTTGTCTTTGATCCAGCACAAATCAAGTCGAACA 3953
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
 Db 3954 GCTAGAGGTGTGCGAGTGCACGAGATCATTTACCTTCTTATCCCTTACTCTCAATG 4013
 QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340
 Db 4014 TCTGAATTCGAAGACGAATCGCTGCTATTTATCCCTTCTGTTATCGTTCAAAACCATGG 4073
 QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
 Db 4074 GTACACAGATTCAGGCGCAGAACCAACCAACCAACCAACCAACCAACCAACCAACCA 4133
 QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
 Db 4134 CCGCAACCTGCGACCAATCTTAAATAGACTCAAATCTCTTGTGTTAGTCAGCTGGTA 4193
 QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyLysSerArgTyrValPheAla 400
 Db 4194 CGAAAAGTTGGGAGGATATGTTATCGAAGAAAAGGATCTCTCTGTTATGCTTTGCG 4253
 QY 401 LysAspLeuProSerSerGluThrValLysAsnLeuGluSerLysGlnGluSer 420

Db 4254 AAAGATTACCATCTGAACATGTTAAATAATCTGAAACCAAGTATTAACAAACAGAGAGT 4313
 Qy 421 ValSerHisThrLeuThrAlaLysGluValAlaProArgAspGlnGluPheTyr 440
 Db 4314 GTTTTACACACATTTAACTGCTTAAAGAAAGAAATGTTGCTCTGTCGACCAAGATTTTAT 4373
 Qy 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
 Db 4374 GATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTGTTTGAATAATTAAGGGTCGT 4433
 Qy 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
 Db 4434 AATCTGATTTCCAAAGCTTAGACAATATTAGAACCTTGAATGATGATCACTAAT 4493
 Qy 481 LysGluLysLeuValAspAspLeuAlaPheLeuAlaProLysHisProGluArg 500
 Db 4494 AAAGAAATTTGGTAGATGATTTATGTCATTTCTAGCACCAATATCCCATCCAGAGCGA 4553
 Qy 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
 Db 4554 CTTGGCAACCAAAATCTCAAAATGAGTATCTGAAGACGAAAGTTCGTATTGCTCAATTA 4613
 Qy 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleLeuSerAsp 540
 Db 4614 GCTGATAAGTATACACGTCAGATGTTACATTTTGTGATGACATGATATATCACTGAT 4673
 Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisThrIleGlyLysAspSer 560
 Db 4674 GAAGGAGATGATATGTAAGCCCTCATATGGCCATAGTCACTGGAATGGAAAGATAGC 4733
 Qy 561 LeuSerAspLysGluLysValAlaAlaGluAlaTyrThrLysGluLysGlyIleLeuPro 580
 Db 4734 CTTTCTGATAGGAAAGATTTGACGCTCAAGCCCTATATCAAGAAAGATGATCTCTACCT 4793
 Qy 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600
 Db 4794 CCATCTCCAGACGAGATGTTAAAGCAATCCCACTGGAGATAGTGAGAGATATTATAC 4853
 Qy 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
 Db 4854 AATCGTGTAAAGGGGAAAAACGAATCCCACTCGTTCATATATGTTGAGCAT 4913
 Qy 621 ThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIle 640
 Db 4914 ACAGTTGAGTTTAAACCGGTAATTTGATTTATCTCATAGGATCAATACCAATAATATT 4973
 Qy 641 LysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660
 Db 4974 AATTTGCTTGGTTTGATGATCACACATACAGCTCCAAATGGCTATATCTTGGAAAGAT 5033
 Qy 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
 Db 5034 TTGTTTTCGACGATTAAGTACTAGTAGAGACACCCCTGACGAAACGCTCCACATTTCTAATGAT 5093
 Qy 681 GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
 Db 5094 GGATGGGCAATGCCAGTGAGCATGTTTAGGCAAGAAAGAACACACAGTGAAGATCCAAAT 5153
 Qy 701 LysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720
 Db 5154 AAGAACTTCAAGCGGATGAAGCCAGTAGAGAAACACCTGCTGAGCCAGAGTCCCT 5213
 Qy 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
 Db 5214 CAAGTAGAGACTGAAAGAAAGTAGAGCCCACTCAAGAAAGCAGAAAGTTTGTTCGAAA 5273
 Qy 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
 Db 5274 GTAACGGATTTCTAGTCTGAAGGCAATGCAACAGAACTCTAGCTGGTTTACGAAATAAT 5333
 Qy 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780

Db 5334 TTGACTCTTCAATATTGGATAACAATAGTATCATGGCAGAACGAGAAAAAATTACTCGC 5393
 Qy 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796
 Db 5394 TTGTTAAAGGAAGTAATCTTTCATCTGTAAGTAAGGAAAAAATAAAC 5441
 RESULT 7
 ABS56454_09/c
 Continuation (10 of 22) of ABS56454 from base 900001 (Streptococcus pneumoniae type 4 str)
 WP Sequence split into 22 fragments LOCUS ABS56454 Accession ABS56454
 WP Fragment Name Begin End
 WP ABS56454_00 1 110000
 WP ABS56454_01 100001 210000
 WP ABS56454_02 200001 310000
 WP ABS56454_03 300001 410000
 WP ABS56454_04 400001 510000
 WP ABS56454_05 500001 610000
 WP ABS56454_06 600001 710000
 WP ABS56454_07 700001 810000
 WP ABS56454_08 800001 910000
 WP ABS56454_09 900001 1010000
 WP ABS56454_10 1000001 1110000
 WP ABS56454_11 1100001 1210000
 WP ABS56454_12 1200001 1310000
 WP ABS56454_13 1300001 1410000
 WP ABS56454_14 1400001 1510000
 WP ABS56454_15 1500001 1610000
 WP ABS56454_16 1600001 1710000
 WP ABS56454_17 1700001 1810000
 WP ABS56454_18 1800001 1910000
 WP ABS56454_19 1900001 2010000
 WP ABS56454_20 2000001 2110000
 WP ABS56454_21 2100001 2162598
 Alignment Scores:
 Pred. No.: 1 98e-280 Length: 110000
 Score: 4183.00 Matches: 795
 Percent Similarity: 99.87% Conservative: 0
 Best Local Similarity: 99.87% Mismatches: 1
 Query Match: 99.95% Indels: 0
 DB: 7 Gaps: 0
 US-09-765-271-56 (1-796) x ABS56454_09 (1-110000)
 Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
 Db 106949 TCTTACGAGTTGGAGCTGTATCAAGCTAGAACGTTAAAGGAAATAATCTGTTTCTAT 106890
 Qy 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
 Db 106889 ATAGATGGAAAAACGACGACGCAAAAAACGAGAAATTTGACTCTGATGAGTTAGCAAG 106830
 Qy 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
 Db 106829 CGTGAAGGATCAATGCTGAGCAATTCGTATCAAGATAACAGACCAAGCTATGCTACT 106770
 Qy 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
 Db 106769 TCACATGGCGACCACTATCATTTATTAATGTTAAGTTCTTATGACGCTATCATCAGT 106710
 Qy 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
 Db 106709 GAAGAATTACTATGAAGATCCAAACTATAAGCTAAAGATGAGGATATTGTTAATGAG 106650
 Qy 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysLysAspAla 120
 Db 106649 GTCAAGGGTGGATATGTTATCAAGGTAGTAGGAAAAATACTATTTTACCTTAAAGGATGCT 106590
 Qy 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
 Db 106589 GCCCAGCGGATAACGTCGCTACAAAGAGGAATCAATCGACAAAAACAAGAGCATAGT 106530
 Qy 141 GlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160

Db	106529	CAACATCGTGAAGGTGGAACTCCAAAGAACGATGCTGTGCTTGCAGCTTCGCAA	106470	Db	105449	CTTGGCAAAACCAAAATCTTCAAATTTAGTATATCTGAAGACGAAGTTCTGTAATTCCTCAATTA	105390
Qy	161	GlyArgTyrThrThrAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr	180	Qy	521	AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp	540
Db	106469	GGACGCTATCTACAGATGATGGTTATATCTTTATGCTTCTGATATCATAGAGTACT	106410	Db	105389	CCTGATTAAGTATACAACGTCAGATGGTTACATTTTGTATGAACATGATATATCAGTGAT	105330
Qy	181	GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu	200	Qy	541	GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer	560
Db	106409	GGTGAATGCTTATATCGTTTCTCATGGAGATCATTAACCATTAATCTCTTAAGATGAGTTA	106350	Db	105329	GAAGCAGATGCATATGTAACGCCTCATATGGCCCATAGTCACTGGATTTGGAAAAAGATASC	105270
Qy	201	SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn	220	Qy	561	LeuSerAspLysGluLysValAlaAlaGluAlaTyrThrLysGluLysGlyLysLeuPro	580
Db	106349	TCAGCTAGCAGTTGGCTGCTGCAGAGCTTCTCTATCTGGTCGAGGAAATCTGTCAAAAT	106290	Db	105269	CTTTCTGATAGGAAAAAGTTGCGACTCAAGCCTATTAAGAAAAAGGATATCTTACT	105210
Qy	221	SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer	240	Qy	581	ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaAlaIleTyr	600
Db	106289	TCAGAAGCTATCGCCGACAAATAGCGATAACACTTCAAGAACAAACCTGGGTACTCTTCT	106230	Db	105209	CCAICTCCAGACGCGAGATGTTAAGCAAAATCCAACCTGGAGATAGTCAGCAGCTATTATAC	105150
Qy	241	ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnSerAsnThrAsnSerGln	260	Qy	601	AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis	620
Db	106229	GTAACCAATCCAGGAACCTACAATTAATACTAACACAGCAACACAGCAACACTAACAGTCAA	106170	Db	105149	AATCGTGTGAAGGGGAAAAACGAATTCACCTCGTTCGACTTCCATATATGTTGAGCAT	105090
Qy	261	AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer	280	Qy	621	ThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIle	640
Db	106169	GCAAGTCAAAAGTAATGACATTGATAGTCTCTTTGAAACAGCTCTCAAAACCTGGCTTTGAGT	106110	Db	105089	ACAGTTGAGGTTAAAAACGTAATTTGATTTATCTCTAAGAGATCATTCACATATATTT	105030
Qy	281	GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr	300	Qy	641	LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp	660
Db	106109	CAACGACATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCAACAAATCAAGTCCGAACA	106050	Db	105029	AAATTTGCTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGAAGAT	104970
Qy	301	AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet	320	Qy	661	LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp	680
Db	106049	GCTAGAGGTGTGCGAGTCCACACCGAGATCATTAACCATTCATCCCTTACTCTCAATG	105990	Db	104969	TTGTTTGGCAGCATTAAGTACTACGTAGAACACACCCCTGACGACGTCACATTCCTAATGAT	104910
Qy	321	SerGluLeuGluGluArgIleAlaArgIleProLeuArgTyrArgSerAsnHisTyr	340	Qy	681	GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn	700
Db	105989	TCTGAATTTGGAAGAACGAATCGTCTGATTAATTCCTCTGTTTATCGTTCAAAACCATTTGG	105930	Db	104909	GGATGGGCAATGCCAGTGAGCATGTGTAGCAAGAAAGACACAGTGAAGATGCCAAT	104850
Qy	341	ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly	360	Qy	701	LysAsnPheLysAlaAspGluProValGluGluThrProAlaGluProGluValPro	720
Db	105929	GTACAGATTTCAAGCCGACAAACCAACAGTCCACACCCGACCTCCGAACTAGTCAGGC	105870	Db	104849	AAGAACTTCAAGCCGATGAAGAGCCAGTAGAGGAAACACCTGTCTGAGCCAGAGTCCCT	104790
Qy	361	ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal	380	Qy	721	GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys	740
Db	105869	CCGCACCTGCACCAAACTTAAATATAGACTCAATTCCTCTTGTGTAGTCAGCTGTA	105810	Db	104789	CAAGTAGAGACTGAAAAAGTAGAACCCCAACTCAAAGAACGACAGAGTTTGTTCGAAA	104730
Qy	381	ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla	400	Qy	741	ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn	760
Db	105809	CGAAAAAGTTGGGAAGGATATGTAATTCGAAGAAAGAGGCAATCTCTCGTTATGCTTTGCG	105750	Db	104729	GTAACGATTTAGTCTGAAGCCCAATGCCACAGAACTCTAGCTGGTTTACGAAATAT	104670
Qy	401	LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer	420	Qy	761	LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla	780
Db	105749	AAAGATTTTACCATCTGAAACTGTATAAAATCTTGAAGCAAGTTATCAAAACAAAGAGAT	105690	Db	104669	TTGACTCTTCAAATTTATGTAACAATAGTATCATGGCAGAGACGAGAAAAATTTACTTGG	104610
Qy	421	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	440	Qy	781	LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn	796
Db	105689	GTITTCACACACTTAACTCTAAAAAGAAATATGTTGCTCTCTCGTCAACAGAAATTTAT	105630	Db	104609	TTGTTAAAGGAAGTAATCTTCTCATCTGTAGTAAGGAAAAATAAAC	104562
Qy	441	AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg	460	Db	104609	TTGTTAAAGGAAGTAATCTTCTCATCTGTAGTAAGGAAAAATAAAC	104562
Db	105629	GATAAAGCATATAATCTGTAACTCAGGGCTCAATAAGCCCTTGTGTAATAATAGGGTCT	105570	Db	104609	TTGTTAAAGGAAGTAATCTTCTCATCTGTAGTAAGGAAAAATAAAC	104562
Qy	461	AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn	480	Db	104609	TTGTTAAAGGAAGTAATCTTCTCATCTGTAGTAAGGAAAAATAAAC	104562
Db	105569	AATCTGATTTCCAGCCCTTAGCAAAATTTATAGAACGCTTGNATGATGATGACTAAT	105510	Db	104609	TTGTTAAAGGAAGTAATCTTCTCATCTGTAGTAAGGAAAAATAAAC	104562
Qy	481	LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg	500	Db	104609	TTGTTAAAGGAAGTAATCTTCTCATCTGTAGTAAGGAAAAATAAAC	104562
Db	105509	AAAGAAAAATTGGTAGATGATTTATTTGGCATTCCTAGCACCATAATCCCATCCAGAGCGA	105450	Db	104609	TTGTTAAAGGAAGTAATCTTCTCATCTGTAGTAAGGAAAAATAAAC	104562
Qy	501	LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu	520	Db	104609	TTGTTAAAGGAAGTAATCTTCTCATCTGTAGTAAGGAAAAATAAAC	104562

RESULT 8

ABS56454_10/c

Continuation (11 of 22) of ABS56454 from base 1000001 (Streptococcus pneumoniae type 4 s

WP Sequence split into 22 fragments LOCUS ABS56454 Accession ABS56454

WP Fragment Name Begin End

WP ABS56454_00 1 110000

WP ABS56454_01 100001 210000

WP ABS56454_02 200001 310000

WP ABS56454_03 300001 410000

WP ABS56454_04 400001 510000

WP ABS56454_05 500001 610000

WP ABS56454_06 600001 710000

WP ABS56454_07 700001 810000

WP ABS56454_08 800001 910000

WP ABS56454_09 900001 1010000

WP	ABSS56454_10	1000001	1110000	
WP	ABSS56454_11	1200001	1210000	
WP	ABSS56454_12	1300001	1310000	
WP	ABSS56454_13	1400001	1410000	
WP	ABSS56454_14	1500001	1510000	
WP	ABSS56454_15	1600001	1610000	
WP	ABSS56454_16	1700001	1710000	
WP	ABSS56454_17	1800001	1810000	
WP	ABSS56454_18	1900001	1910000	
WP	ABSS56454_19	2000001	2010000	
WP	ABSS56454_20	2100001	2110000	
WP	ABSS56454_21	2200001	2210000	

Alignment Scores:

Pred. No.:	1.98e-280	Length:	110000
Score:	4163.00	Matches:	795
Percent Similarity:	99.87%	Conservative:	0
Best Local Similarity:	99.87%	Mismatches:	1
Query Match:	99.95%	Indels:	0
DB:	7	Gaps:	0

US-09-765-271-56 (1-796) x ABSS56454_10 (1-110000)

Qy	1	SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVallysGluAsnAsnArgValSerTyr	20
Db	6949	TCITTACAGTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTGTGTTCCAT	6890
Qy	21	IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys	40
Db	6889	ATAGATGGAAAAACAGCGACGCAAAAACGGAGATTGTGACTCTGTATAGGTTAGCAAG	6830
Qy	41	ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr	60
Db	6829	CGTGAAGGAATCAATGCTGAGCAAAATCGTTCATCAAGATAACAGACCAAGCTATGTCACT	6770
Qy	61	SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleLeuSer	80
Db	6769	TCATATGGCGACCACTATCATATTATTAACATGGTAAGGTTCCITTTATGACGCTATCATCAGT	6710
Qy	81	GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu	100
Db	6709	GAGAAATTACTCATGAAGATCCAACTATAGCTAAAGATGAGGATATGTTAATGAG	6650
Qy	101	ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla	120
Db	6649	GTCAAGGGTGGATATGTTATCAAGTAGATGGAATAATACTATGTTACCTTAAAGGATGCT	6590
Qy	121	AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer	140
Db	6589	GCCACGCGGATTAACGTCGCTACAAAAGAGGAATAATCAATCGCAAAAACAAAGACATAGT	6530
Qy	141	GlnHisArgGluGlyGlyThrProArgAsnAspGlyValaValAlaLeuAlaArgSerGln	160
Db	6529	CAACATCGTGAAGTGGAACTCCAAGNAACGATGGTGTCTGTGCCCTTGGCACGTTCCGAA	6470
Qy	161	GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr	180
Db	6469	GGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTCATATCATAGAGGATACT	6410
Qy	181	GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu	200
Db	6409	GGTGATGCTTATATCGTTCCTCATGGAGATCATTTACCATTTACATTCTTAAGAATGAGTTA	6350
Qy	201	SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn	220
Db	6349	TCAGCTAGCGAGTTGGCTGCTGCAGAGCCCTTCTTATCTGGTCGAGAAATCTGTCAAAAT	6290
Qy	221	SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer	240
Db	6289	TCAAGAACCCTATCCCGACAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTCT	6230
Qy	241	ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln	260

QY 621 ThrValGluValLysAsnGlyAsnLeuLeuLeuProHisLysAspHisLysHisAsnIle 640
 Db 5089 ACAGTTGAGGTAAACACGGFAATTGATTATCTCATAGGATCATTAACCAATATT 5030
 QY 641 LysPheAlaTTPheAspHisThrLysAlaProAsnGlyTyrThrLeuGluAsp 660
 Db 5029 AAATTTGCTTGGTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTGGAGAT 4970
 QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
 Db 4969 TTGTTTGGCGACGATTAGTACTACGTAGACACACCTGACGACGTCACATCTTAATGAT 4910
 QY 681 GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysAspHisSerGluAspProAsn 700
 Db 4909 GGATGGGCAATGGCAGTGAGCATGTGTGGCAAGAAAGACCAAGTGAAGATCCCAAT 4850
 QY 701 LysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720
 Db 4849 AAGAACTTCAAGCGGATGAGAGCCAGTACAGGAAACACCTGCTGAGCCAGAGTCCCT 4790
 QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlalys 740
 Db 4789 CAAGTAGAGACTGAAAGAGTAGAGCCCACTCAAGAGAGAGAGTTTGGTTGCGAAA 4730
 QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
 Db 4729 GTAAAGGATTTAGTCTGAAAGCCCAATGCAACAGAACTCTAGCTGGTTTACGAATTAAT 4670
 QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780
 Db 4669 TTGACTCTTCAATTTATGATAACAAATAGTATCATGGCAGAGCAGAGAAATTAATTCTGCG 4610
 QY 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796
 Db 4609 TTGTTAAAGGAAGTAATCTTCTCATCTGTAAGTAAGGAAAAATAAAC 4562

RESULT 9
 AAA65731
 ID AAA65731 standard; DNA; 2523 BP.
 AC
 AC AAA65731;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Streptococcus pneumoniae BVH-11 gene SEQ ID NO:3.
 XX
 DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal; ds.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA001219.
 XX
 PR 23-DEC-1998; 98US-0113800P.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX
 DR WPI; 2000-452397/39.
 XX
 DR P-PSDB; AAB12716.
 XX
 PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia.
 XX
 PS Example 2; Fig 3; 106pp; English.
 XX

CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,
 CC may be used for the recombinant production of the proteins they encode.
 CC The protein antigens may then be used as vaccines for the prevention and
 CC treatment of Streptococcal infections in mammals (especially humans)
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
 CC pneumonia. The present sequence encodes the S. pneumoniae BVH-11 protein
 CC antigen
 XX
 SQ Sequence 2523 BP; 879 A; 523 C; 526 G; 595 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 3.81e-216 Length: 2523
 Score: 3218.00 Matches: 616
 Percent Similarity: 82.95% Conservative: 65
 Best Local Similarity: 75.03% Mismatches: 102
 Query Match: 77.26% Indels: 38
 DB: 3 Gaps: 3
 US-09-765-271-56 (1-796) x AAA65731 (1-2523)
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
 Db 61 GCTTATGAACTAGGTTTGCATCAAGCTCAACTGTAAGAAATAATTCGTGTTTCCCTAT 120
 QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
 Db 121 ATAGTGAAGAAACCAAGCGACGCAAAACGAGGAAATTTGACTCTCTGATGAGGTAGCAAG 180
 QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
 Db 181 CGTGAAGGAATCAACGCCGAAACAAATCGTCATCAAGATTACGATCAAGTTTATGTGACC 240
 QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
 Db 241 TCTCATGGAGACCATTTATCTATCTATTAATGCGAAGTCCCTTATGATGCCCATCATCACT 300
 QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
 Db 301 GAAGACTCTCTCATGAAGATCCGAAATTTATCAGTTGAAGATTACAGATTTGTCATGAA 360
 QY 101 ValLysGlyTyrValIleLysValAspGlyLysTyrTyrValTyrValTyrLeuLysAspAla 120
 Db 361 ATCAAGGCTGTTATGTCTAATGAAGTAAACGGTAAATATCTATGTTTACCTTAAGGATGCA 420
 QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnLysHisSer 140
 Db 421 GCTCATGGGATAATGTCGTACAAAAGAAATAATCAATCGCAAAAACAGAACATAGT 480
 QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
 Db 481 CAGCATCTGAGGAGGAGCTTCAGCAAAACGATGTCGGTAGCCTTTGACGTTTACACAG 540
 QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
 Db 541 GGACGCTACACCCACAGATGATGTTTATATCTTCAATGTCATCTGATATCATTCAGATACG 600
 QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
 Db 601 GCGGATGCTATATCTTCCTCATGGAGATCATTAACCATTAATCTTCTTAAGATGAGTTA 660
 QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
 Db 661 TCAGCTAGCGAGTTGGCTGTCGAGAAGCCTTCTCTATCTGTCGGGAAAATCTGTCAAT 720
 QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
 Db 721 TTAAGAACCTATCGCCGACAAAATAAGCGATAACACTCCAAGAACAAACTGGGTACTCTCT 780
 QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
 Db 781 GTAAGCAATCCAGGAACCTACAAATACTAACACAGCAACACAGCAACACTAACAGTCAA 840

QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
 DB 841 GCAAGTCAAGTAATGACATGATAGTCTCTTGAACAGCTCTCAAAAGTCCTTTGAGT 900
 QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
 DB 901 CAACGCCATGTAGATCTGATGGCTTATTTTCGACCCAGCGCAATCAAGTCGAACC 960
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
 DB 961 GCCAGAGTGTAGTGTCTCCCTCATGGTAAACCATTAACCACTTTATCCCTTATGAACAAATG 1020
 QY 321 SerGluLeuGluArgGlnIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340
 DB 1021 TCTGAATGGAAAACGAATGCTCGTATTAATCCCTTCCTTCGTTTCGTTCAACCAATGG 1080
 QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
 DB 1081 GTACCAAGTTCAGACCAAGAAACCAAGTCCACACCGACTCCAGAACCTTAGTCCAAGT 1140
 QY 361 ProGlnProAlaProAsnLeuLys-----IleAspSerAsnSerSer 374
 DB 1141 CCGCAACTGCACCAATCTCTCAACAGCTCCAAGCAATCCAATTGAT-----GAGAAA 1194
 QY 375 LeuValSerGlnLeuValArgLysValGlyGlyTyrValPheGluGluLysGlyIle 394
 DB 1195 TTGTCACAAAGAGCTGTTCCGAAGTAGTGGTATGTTATGCTTTGAGGAGAATGGAGTT 1254
 QY 395 SerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuSerLys 414
 DB 1255 TCTCGTTATATCCAGCAAGAAATCTTTCCAGCAAGAACAGCAGCAGCAATGTATGACAAA 1314
 QY 415 LeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaPro 434
 DB 1315 CTGGCCAGCAGGAAGTTTATCTCAAGTAGGCTATGACTTACTAGCAAGAACTGACCTCCCATCT 1374
 QY 435 ArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeu 454
 DB 1375 AGTGATCGAGATTTTACATAGGCTTATGACTTACTAGCAAGAACTGACCTCCCATCT 1434
 QY 455 Phe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeu 474
 DB 1435 CTTGATAATAAAGTCGACAAAGTTGATTTTGAGGCTTTGGATAACCTGTTGGAAACGACTC 1494
 QY 475 AsnAspGluSerThrAsnLysGluLeuValAspAspLeuLeuAlaPheLeuAlaPro 494
 DB 1495 AAGGATGCTCAAGTGATAAGTCAAGTTAGTGGATGATATTCTTGCTTTCTTAGCTCCG 1554
 QY 495 IleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGlu 514
 DB 1555 ATTCGTCATCCAGAACGTTTAGGAAAACCAATGCGCAATTAACCTACACTGATGATGAG 1614
 QY 515 ValArgIleAlaGlnLeuAlaAspLysTyrThrSerAspGlyTyrIlePheAspGlu 534
 DB 1615 ATTCAGTAGTCCCAAGTTGGCAGGCAAGTACACACAGAACGGTTATATCTTTGATCTCT 1674
 QY 535 HisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHis 554
 DB 1675 CGTGATATAACCAAGTATGAGGGGATGCTTGTAACTCCACATATGACCCCATAGCCAC 1734
 QY 555 TrpIleGlyLysAspSerLeuSerAspLysValAlaAlaGlnAlaTyrThrLys 574
 DB 1735 TGGATTAATAAAGATAGTTTGTCTGAAGCTGAGAGCGCGCAGCCAGGCTTATGCTAAA 1794
 QY 575 GluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGlyAsp 594
 DB 1795 GAGAAAAGTTTGAACCTCCTCCAGACACCATCAGGATTCAGAAATCTAGAGGCAAAA 1854
 QY 595 SerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArgLeu 614
 DB 1855 GGAAGCAGAGCTATCTACAAACCGCGTGAAGAGCAGCTAAGAAAGTGCCACTGTGCTATG 1914

QY 615 ProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHisLys 634
 DB 1915 CTTTACCAATCTTCAATATATCTGTAGAGTCAAAAACGGTAGTTTAATCATACCTCATAT 1974
 QY 635 AspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsn 654
 DB 1975 GACCAATACCAATACATCAAAATTTGAGTGGTTTTCAGCAAGCGCTTTATGAGCGCACTAAG 2034
 QY 655 GlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGlu 674
 DB 2035 GGTATATCTCTTGAGGATCTTTTGGCGACTCTCAAGTACTATGTCGAACATCCCAACGAA 2094
 QY 675 ArgProHisSerAsnAspGlyTyrPheAsnAlaSerGluHisValLeuGlyLysLysAsp 694
 DB 2095 CGTCCGCAATTCAGATAATGGTTTGGTAAACGCTAGCGACCATGTTCAAAGAAACAAAAAT 2154
 QY 695 HisSerGluAspProAsnLysAsnPheLysAlaAsp----- 706
 DB 2155 GGTCAAGCTGATACCAATCAAAACGAAACCAAGCGGAGAAACCTCAGACAGAAAAA 2214
 QY 706 ----- 706
 DB 2215 CCTCAGGAAGAAACCCCTCGAGAAAGAGAAACCAAAACGAGAGAGTCTCCAAAA 2274
 QY 707 -----GluGluProValGluGluThrProAlaGluProGluValProGlnValGluThr 724
 DB 2275 CCAACAGAGGAACCCAGAGGAAGAAATCACCAGAGGAATCAGAAAGAACCTCAGGTCGAGACT 2334
 QY 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuAlaLysValThrAspSer 744
 DB 2335 GAAAGGTTGAAGAAACAACTCAGAGAGGCTGAGATTTACTTGGAAAAAATATTTTACTATTGGC 2394
 QY 745 SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGln 764
 DB 2395 ATTCAAGTCCCAATGCGAAAGAGACTCTCACAGGATTAATAATAATTTTACTATTGGC 2454
 QY 765 IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGly 784
 DB 2455 ACCAGGACCAACAATACTATTATGGCAGAAAGCTGAAAAACATTTGGCTTTTATAAGGAG 2514
 QY 785 Ser 785
 DB 2515 AGT 2517
 RESULT 10
 ID AAA65736 standard; DNA; 2647 BP.
 AC AAA65736;
 DT 21-NOV-2000 (first entry)
 DE Streptococcus pneumoniae BVH-11 gene SEQ ID NO:12.
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal; ds.
 OS Streptococcus pneumoniae.
 PN WO200039299-A2.
 PD 06-JUL-2000.
 PF 20-DEC-1999; 99WO-CA001218.
 PR 23-DEC-1998; 98US-0113800P.
 PA (BIOC-) BIOCHEM PHARMA INC.
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 DR WPI; 2000-452397/39.

XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteremia and/or pneumonia.
 XX Example 6; Fig 15; 106pp; English.

CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,
 CC may be used for the recombinant production of the proteins they encode.
 CC The protein antigens may then be used as vaccines for the prevention and
 CC treatment of Streptococcal infections in mammals (especially humans)
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
 CC pneumonia. The present sequence encodes the S. pneumoniae BVH-11 protein
 CC antigen

XX Sequence 2647 BP; 934 A; 538 C; 556 G; 619 T; 0 U; 0 Other;

Alignment Scores: 4.04e-216 Length: 2647
 Pred. No.: 3218.00 Matches: 616
 Score: 82.95% Conservative: 65
 Percent Similarity: 75.03% Mismatches: 102
 Best Local Similarity: 77.26% Indels: 38
 Query Match: 3 Gaps: 3
 DB:

US-09-765-271-56 (1-796) x AAA65736 (1-2647)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
 DB 105 GCTTATGAACATAGGTTTGCATCAAGCTCAACATGTAAGAAATAATCGTGTTCCTAT 164
 QY 21 IleAspGlyLysGluAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
 DB 165 ATAGATGGAACACAGCGCAGCAAAACCGAGAAATTTGACTCCGTGAGGTAGCAAG 224
 QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
 DB 225 CQTGAAGGAATCAACCGCAACATCGTCATCAAGATTACGATCAAGTTATGTGACC 284
 QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleLeuSer 80
 DB 285 TCTCATGAGACCAATATCATTTACTATATATGCGCAAGTCCCTTATCATGCCATCATCAGT 344
 QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
 DB 345 GAAGAGCTCTCATGAAGATCCGAATATATCAATGAGGATTCAGACATTCATCAATGAA 404
 QY 101 ValLysGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysLysAspAla 120
 DB 405 ATCAAGGGTGGTTATGTCATTAAGGTAAACGGTAAATATCTATGTTTACCTTAAGGATGCA 464
 QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
 DB 465 GCTCATGCGGATAATGTCGTACAAAGAAAGAAATCAATCGCGCAAAACCAAGAACATAGT 524
 QY 141 GlnHisArgGluGlyTyrProArgAsnAspGlyValAlaValAlaLeuAlaArgSerGln 160
 DB 525 CAGCATCGGAAGGAGGACTTCAGCAACAGATGGCGGTAGCCCTTTCAGCTTCACAG 584
 QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
 DB 585 GGAACGCTACACACAGATGATGTTTATCTTCAATGATCTGATATCATCGAAGATACG 644
 QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
 DB 645 GCGCATGCGCTATATCTCTCATGAGATCAATACCATCATCTTCAAGATGAGTTA 704
 QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
 DB 705 TCAGCTAGCGAGTTGGCTGCTGCAGAACGCTTCTCTATCTGGTCGGGAAATCTGTCAAT 764
 QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240

DB 765 TTAAGAACTATCGCCGACAAAATAGCGATAAACAACCTCCAAAGCAAAATCGGTACCTTCT 824
 QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
 DB 825 GTAAGCAATCAGGAACCTACAAATACTAACCAACAGCAACCAACAGCAACCAACAGTCAA 884
 QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
 DB 885 GCAAGTCAAGTAAATGACATGATGATGCTCTTGAACAGCTCTACAAACCTGCTTGGAT 944
 QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
 DB 945 CAACGCCATAGTAATCTGATGGCTTATTTTCGACCCAGCGCAAAATCACAAGTCGAACC 1004
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
 DB 1005 GCCAGAGGTAGCTGCTCCCTCATGTTAAACCATTACCACCTTTATCCCTTATGAACAAATG 1064
 QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340
 DB 1065 TCTGAAATGGAACAAACGAATTCCTGCTGATTTATCCCTTCGTTCGTTCGTTCAACACCTGG 1124
 QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
 DB 1125 GTACCAGATTCAAGACCAAGAACCAAGTCCCAACCGACTCCCAAGAACCTAGTCCAAAT 1184
 QY 361 ProGlnProAlaProAsnLeuLys-----IleAspSerAsnSerSer 374
 DB 1185 CGCAACCTGCACCAAAATCCTCAACCGACTCCAGCAATCCCAATTCAT-----GAGAAA 1238
 QY 375 LeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIle 394
 DB 1239 TTGGTCAAGAGCTGTTTCGAAAGTAGCGCATGCTTATGCTTTGAGGAGAATCGAGTT 1298
 QY 395 SerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLys 414
 DB 1299 TCTCGTTATATCCACCAAGAAATCTTTACAGCAGAAACAGCAGCAGGAGCATTTGATGACAAA 1358
 QY 415 LeuSerLysGlnGluSerValSerHisIleThrLeuThrAlaLysLysGluAsnValAlaPro 434
 DB 1359 CTGGCCAGCAGGAAAGTTTCTCTAAGCTAGAGCTAGAGAGCTAGAGAACTGACCTCCCATCT 1418
 QY 435 ArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeu 454
 DB 1419 AGTGATCGAGAATTTTACAATAAGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTA 1478
 QY 455 Phe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuGluArgLeu 474
 DB 1479 CTTGATAATAAAGGTGCAAGTTGATTTTGGGCTTTGGATAAAGCTGTTGGAACGACTC 1538
 QY 475 AsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuAlaPheLeuAlaPro 494
 DB 1539 AAGGATGCTCAAGTGATAAAGTCAAGTTAGTGGATGATATCTTTCCTTCTTAGCTCCG 1598
 QY 495 IleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGlu 514
 DB 1599 ATTGCTATCCAGAACGTTTAGGAAAAACCAATGGCAAAATTTACCTACATGATGATGAG 1658
 QY 515 ValArgIleAlaGlnLeuAlaAspLysTyrThrSerAspGlyTyrIlePheAspGlu 534
 DB 1659 ATTCAAGTAGCCAAAGTTGCGCAGGCAAGTACACAAAGAGACGGTATATCTTTGATCCT 1718
 QY 535 HisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHis 554
 DB 1719 CGTGATATACCAAGTATGAGGGGATGCTTATGTAACCTCCACATATGATGCCATGCCAC 1778
 QY 555 TrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaGlnAlaTyrThrLys 574
 DB 1779 TGGAAATTAAGAGATAGTTTGTCTGAAGCTGAGAGAGCGCGCAGCTTATGCTATAA 1838
 QY 575 GluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGlyAsp 594

Db 1839 GAGAAAGGTTTGACCCCTCTTCGACAGACCATCAGGATTTCAGGAAATCTAGGCAAAA 1898
 QY 595 SerAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArgLeu 614
 Db 1899 GGAGCAGAGCTATCTCAACCGCGTGAAGCAGCTAAGAAGGTGCCACTTGATCGTATG 1958
 QY 615 ProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHisLys 634
 Db 1959 CCTTACATCTTCATATAGTCTAGAGTCAAAACGGTAGTTTAATCATCTCATTTAT 2018
 QY 635 AspHisTyrHisAsnIleIlePheAlaTrpPheAspHisThrTyrLysAlaProAsn 654
 Db 2019 GACCATTAACCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 2078
 QY 655 GlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrValGluHisProAspGlu 674
 Db 2079 GGGTATCTCTTGAGGATCTTTTGGCGACTCTCAAGTACTATGTCGAACATCCCAACGAA 2138
 QY 675 ArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAsp 694
 Db 2139 CGTCGGCATTCAGATAATGTTTGGTAAACGCTAGCGACCATGTTCAAGAAGAAAT 2198
 QY 695 HisSerGluAspProAsnLysAsnPhelLysAlaAsp 706
 Db 2199 GGTCAAGCTGTATACCAATCAACCGGAAACCAAGCGAGGAGAAACCTTCAGACAGAAAA 2258
 QY 706 2259 CCTGAGGAAGAAACCTTCGAGAGAGAAACCAAGCGAGGAGAAACCTTCAGAGTCTCCAAAA 2318
 QY 707 2319 CCAACAGAGAAACCAAGAGAGAAATCACCAGAGGAAATCAGAGACCTCAGTCGAGACT 2378
 QY 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer 744
 Db 2379 GAAAGGTTGAAGAAACCAAGAGAGGCTGAGAGGCTGAGATTCTTGGAAAAATCCAGGATCCA 2438
 QY 745 SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGln 764
 Db 2439 ATTATCAAGTCCATGCAAGAGACTCTCACAGGATTAATAATATTACTATTGTCG 2498
 QY 765 IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGly 784
 Db 2499 ACCAGCAGACAACTATCTATTATGCGAGAGCTGAGAAACATTTGGCTTTATTAAAGAG 2558
 QY 785 Ser 785
 Db 2559 AGT 2561
 RESULT 11
 ID ABK15103
 XX
 AC ABK15103;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE DNA encoding Streptococcus pneumoniae BVH-11.
 XX
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;
 KW streptococcal bacterial infection; gene; ds.
 XX
 OS Streptococcus pneumoniae.
 XX
 PH Location/Qualifiers
 FT CDS 45..2567
 FT /tag= a
 FT /product= "BVH-11"
 FT /note= "The gene is flanked by sequences from the vector
 FT SP64, no information on which is given in the
 FT specification"

PN WO200198334-A2.
 XX
 PD 27-DEC-2001.
 XX
 PF 19-JUN-2001; 2001WO-CA000908.
 XX
 PR 20-JUN-2000; 2000US-0212683P.
 XX
 PA (SHIR-) SHIRE BIOCHEM INC.
 XX
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX
 DR WPI; 2002-122272/16.
 DR P-FSDB; AAU75933.
 XX
 PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing
 PT polypeptides, useful as vaccine components for treating or preventing
 PT streptococcal infections such as otitis media, meningitis, and
 PT bacteremia.
 XX
 PS Disclosure; Fig 4; 113pp; English.
 XX
 CC The invention describes an isolated polypeptide (I) with 70-90% identity
 CC to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or
 CC BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
 CC Staphylococcus aureus) in an individual susceptible to the infection. A
 CC polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence encodes the
 CC Streptococcus pneumoniae protein BVH-11, used to create the antigenic
 CC peptides described in the method of the invention
 XX
 SQ Sequence 2647 BP; 934 A; 538 C; 556 G; 619 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 4,04e-216 Length: 2647
 Score: 3218.00 Matches: 616
 Percent Similarity: 82.95% Conservative: 65
 Best Local Similarity: 75.03% Mismatches: 102
 Query Match: 77.26% Indels: 38
 DB: 6 Gaps: 3
 US-09-765-271-56 (1-796) x ABK15103 (1-2647)
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
 Db 105 GCTTATGAACACTAGGTTTGCATCAAGCTCAAACTGTAAAGAAAAATAATCGTGTTCCTAT 164
 QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
 Db 165 ATAGATGGAAAAACAAGCAGCGAAAAACGGAAATTTGACTCCTGATGAGGTAGCAAG 224
 QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
 Db 225 CGTGAAGGAATCAACGCCGCAACAAATCGTCATCAAGATTACGATCAGGTTATGTGACC 284
 QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
 Db 285 TCTCATGGAGACCATTTATCATCTACTATAATGCGCAAGGTCCCTTATGATGCCATCATCAGT 344
 QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100

Db 345 GAAGAGCTCCTCATGAAGATCCGAATTTATCAGTTGAAGATTTCAGACATTTGTCATGA 404
Qy 101 VallysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db 405 ATCAAGGGTGTGTATGTCATTAAGGTAACGGTAATAATATGTTTACCTTTAAGGATGCA 464
Qy 121 AlahisAlaAspAsnValArgThrLysGluGluLeuLeuAsnArgGlnLysGlnGluHisSer 140
Db 465 GCTCATGCGGATATGTCGTCACAAAGAGAAATCAATCGGCAAAAACAGACATAGT 524
Qy 141 GlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
Db 525 CAGCATCGTGAAGGAGGACTTCAGAAACGATGTCGGTAGCTTTGCGACGTTTCACAG 584
Qy 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db 585 GGACGCTACACACAGATGATGGTTATATCTTCATGTGATCTGATATCATCGAAGATACG 644
Qy 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db 645 GCGCATGCTATATGTTCTCATGAGATCATTTACCATTACATTCCTTAAGAATGAGTTA 704
Qy 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db 705 TCAGCTAGCGAGTTGGCTGCTGCAGAGCCTTCTCTATCTGTCGGGAAAAATCTGTCAAAT 764
Qy 221 SerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
Db 765 TTAAGAACCTATCGCGCAAAATAGCGATAACCTCCAAAGAACAACTGGGTACTCTTCT 824
Qy 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db 825 GTPACCAATCCAGAACTACAATATACACACAGCAACACAGCAACACATTAACAGTCAA 884
Qy 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
Db 885 GCAAGTCAAAAGTAATGACATTCATGATGCTCTTTGAAACAGCTCTACAAATGCTTTGAGT 944
Qy 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db 945 CAACGCCATGTGAATCTGATGGCTTATTTTCGACCCAGCGCAATCAAAATCAAACTG 1004
Qy 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db 1005 GCACAGGTGTAGTCTCTCATGTATACCATTAACCATTTATCCCTTATGAACAAATG 1064
Qy 321 SerGluLeuGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340
Db 1065 TCTGAATTGGAACGAATGCTCGTATTTATTCCTTCGTTATCGTTTCAAAACCATTTGG 1124
Qy 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db 1125 GTACAGATTCAGACCAAGAACCAACAGTCCACACAGCTCCAGAACCTAGTCCAAGT 1184
Qy 361 ProGlnProAlaProAsnLeuLys-----IleAspSerAsnSerSer 374
Db 1185 CCGCAACCTGCACCAATCTCAACAGCTCCAAAGCAATCCAAATTCAT-----GAGAAA 1238
Qy 375 LeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIle 394
Db 1239 TTGGTCAAAAGAGCTGTTCCGAAAAGTAGCGATGTTATGTCCTTTGAGGAGAAATGGAGTT 1298
Qy 395 SerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLys 414
Db 1299 TCTCGTTATATCCAGCAAGAACTTTTCAGCGAAACAGCAGCAGCGATTCATGATGACAA 1358
Qy 415 LeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaPro 434
Db 1359 CTGCGCAAGCAGCAAGATTTATCTCATAGCTAGGAGCTAAAGAAAACTGACCTCCCATCT 1418
Qy 435 ArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeu 454
Db 1419 AGTGATCGAGAAATTTTACAATAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTA 1478

Qy 455 Phe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeu 474
Db 1479 CTTGATAATAAGGTCGACAAAGTTGATTTTGAGGCTTTGGATAAAGCTTTGGAACGACTC 1538
Qy 475 AsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaPro 494
Db 1539 AAGGATGCTCAAGTGATAAAGTCAAGTTAGTGATGATATTTCTTGGCTTTTACGCTCCG 1598
Qy 495 IleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGlu 514
Db 1599 ATTCTCATCCAGAACGTTTAGGAAAACCAATTCGCCAAATACCTACACATGATGATGAG 1658
Qy 515 ValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGlu 534
Db 1659 ATTCAAGTAGCAAGTTGGCAGCAAGTACACACAGAAAGACGGTTATATCTTTGATCCT 1718
Qy 535 HisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHis 554
Db 1719 CGTGATATAACGAGTATGAGGGGATGCTTATGTAACCTCACATATGACCATGATGCCAC 1778
Qy 555 TrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLys 574
Db 1779 TGGATTAAAAAGATAGTTTGTCTGAAGCTGAGAGCGCGCAGCCAGCTTATGCTAAA 1838
Qy 575 GluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGlyAsp 594
Db 1839 GAGAAAGGTTTGACCCCTCTTCGACAGACCATCAGGATTTCAGGAATAATCTAGGCAAAA 1898
Qy 595 SerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArgLeu 614
Db 1899 GGAGCAAGAGCTATCTACACCCGCGTGAAGCAGCTAAGAAGTGCCACTTCGATCGATG 1958
Qy 615 ProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHisLys 634
Db 1959 CCTTCAATCTTCAATATCTGTAGAAGTCAAAACGCTAGTTTAAATCATACCTCATTTAT 2018
Qy 635 AspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsn 654
Db 2019 GACCATTAACATAACATCAAAATTTGAGTGGTTTGACGAAGGCTTTATGAGCGACCTAAG 2078
Qy 655 GlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGlu 674
Db 2079 GGGTATACCTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGAAACATCAACAGAA 2138
Qy 675 ArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAsp 694
Db 2139 CGTCCGATTCAGATAATGGTTTGGTAACTAGCGCTAGCGACATGTTCAAGAAACAAAT 2198
Qy 695 HisSerGluAspProAsnLysAsnPheLysAlaAsp----- 706
Db 2199 GGTCAAGCTGATACCAATCAAAACGGAACCAAGCGAGGAGAAACCTCAGACAGAAAAA 2258
Qy 706 ----- 706
Db 2259 CCTGAGGAAGAACCCCTCGAAGAGAGAAACCAAGCGAGAAACAGAGTCTCCAAA 2318
Qy 707 -----GluGluProValGluGluThrProAlaGluProGluValProGlnValGluThr 724
Db 2319 CCAACAGAGAACCGAAGAAATCCACAGAGGATCAGAGAACCTCAGGTCGAGACT 2378
Qy 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer 744
Db 2379 GAAAGGTTGAAGAAAACTGAGAGGCTGAAGATTTTACTTTGGAAAAATCCAGGATCCA 2438
Qy 745 SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGln 764
Db 2439 ATTATCAAGTCCATGCCAAAGACTCTCACAGGATTAATAAATAATTTACTATTGTC 2498
Qy 765 IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGly 784
Db 2499 ACCCAGGACAACTACTATTATGCGAAGCTGAAAAAATCTATTGGCTTTTATTAAGGAG 2558

QY 785 Ser 785
 Db 2559 AGT 2561
 RESULT 12
 ID AAA08557 standard; DNA; 2478 BP.
 XX AAA08557;
 AC AAA08557;
 XX 19-JUL-2000 (first entry)
 XX S. pneumoniae 92 kDa human C3-degrading protein coding sequence.
 DE Human C3-degrading protein; 92 kDa; immunostimulatory; vaccine;
 KW inhibitor; inflammation; organ rejection; xenotransplantation; ss.
 KW Streptococcus pneumoniae.
 OS Streptococcus pneumoniae.
 XX WO200017370-A1.
 PN 30-MAR-2000.
 XX 24-SEP-1999; 99WO-US022362.
 XX 24-SEP-1998; 98US-0101736P.
 PR 31-MAR-1999; 99US-00283094.
 XX (MINU) UNIV MINNESOTA.
 PA (AMCY) AMERICAN CYANAMID CO.
 XX Hostetter MK, Finkel DJ, Cheng Q, Green BA, Masi AW;
 XX WPI; 2000-283594/24.
 DR P-PSDB; AAY91939.
 XX Isolated polypeptide is used to stimulate immune system and immunize or
 PT treat a mammalian subject against Streptococcus pneumoniae infection or
 PT colonization.
 XX Claim 1; Page 55-57; 63pp; English.

CC The present sequence, isolated from Streptococcus pneumoniae, encodes a
 CC human C3-degrading protein (see AAY91939) of about 92 kDa. This sequence
 CC may encompass a smaller 20 kDa polypeptide coding sequence (AAA08556).
 CC also having human C3-degrading activity. The DNA sequences can be used
 CC for producing an immune response to Streptococcus pneumoniae in a mammal.
 CC Antibodies against the proteins can be used to inhibit S. pneumoniae-
 CC mediated C3 degradation. C3-mediated inflammation and rejection in
 CC xenotransplantation can be inhibited by expressing the nucleic acid
 CC sequences on the surface of an organ of an animal. In particular, the
 CC polypeptides are useful for stimulating the immune system and are
 CC effective to immunize or treat a mammalian subject against Streptococcus
 CC pneumoniae infection or colonization
 XX SQ Sequence 2478 BP; 837 A; 510 C; 535 G; 596 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.09e-187 Length: 2478
 Score: 2812.00 Matches: 552
 Percent Similarity: 77.10% Conservative: 81
 Best Local Similarity: 67.24% Mismatches: 136
 Query Match: 67.52% Indels: 52
 DB: 3 Gaps: 8

US-09-765-271-56 (1-796) x AAA08557 (1-2478)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19
 Db 61 TCCTATGAACCTTGGTCTGTCACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCAGGTTTCT 120
 QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39

Db 121 TATATAGATGGTGATCAGGCTGGTCAAAAGCGCAGAAAACCTTGACACCAGATGAAGTCACT 180
 QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
 Db 181 AAGAGGGAGGGGATCAACGCCGACAAATCGTCAATCAAGATTACGGATCAAGGTTATGTG 240
 QY 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79
 Db 241 ACCTCTCATGGAGACCAATTATCACTATTAATGCAAGGTCCTCTTATGATGCCATCATC 300
 QY 80 SerGluGluLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
 Db 301 AGTGAAGAGCTCTCATGAAGATCCGAATATCAGTTGAAGATTGACATTTGTCAT 360
 QY 100 GluValLysGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysLeuAsp 119
 Db 361 GAAATCAAGGGTGGTTATGTTATCAAGGTAGATGAAAATACTATGTTTACCTTANGAT 420
 QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluLuleAsnArgGlnLysGlnGluHis 139
 Db 421 GCAGCTCATGGGATAATATTCGACAAAAGAGATTAAACGTCAGAACGAGAACAC 480
 QY 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
 Db 481 AGTCATAATCACGGGGGTGGTTCT-----AACGATCAAGCAGTAGTTGACGCCAGGCC 534
 QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
 Db 535 CAAGGACGCTATACACGGATGATGTTATATCTTCAATCATCTGATATCATTTAGGAC 594
 QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
 Db 595 ACGGGTGATGCTTATATCGTTCTCCACGGCGACCATACCATTCATTCCTCAAGAATGAG 654
 QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
 Db 655 TTATCAGCTAGCAGTTAGTCTGTGAGAACGCTTATGGAAATGGG----- 699
 QY 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239
 Db 700 -----AAGCAGGGATCTCGCTCTTCTTCAAGTTCTAGTTATATAATGCA 741
 QY 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
 Db 742 AATCCAGCTCAACCAAGATTGTCAGAGAACCAACATCTGACTGTCTCACTCAACTTATCAT 801
 QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLysGlnLeuTyrLysLeuProLeu 279
 Db 802 CAA--AATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTGTATGCTAAACCCCTTA 858
 QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
 Db 859 TCAGAACGGCATGTGAATCTGATGGCTTATTTTCGACCCAGCGCAAAATCACAAGTCGA 918
 QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
 Db 919 ACCGCCAGAGGTAGTCTGCTCCCTCATGTGAACCATTTACCATTTATCCCTTATGAACAA 978
 QY 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
 Db 979 ATGCTGTGAATTGGAAAAAGAAATGCTGCTATATTATTCCTCTGTTATGCTTCAAAACCAT 1038
 QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
 Db 1039 TGGGTACCAAGATTCAAGACCAAGAACCAAGTCCACATCGACTCCGGAACCTAGTCCA 1098
 QY 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373
 Db 1099 AGTCGCAACCTGCAACAAATCTTCAACAGCTCCAAAGCAATCCAAATTTGAT-----GAG 1152
 QY 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393
 Db 1153 AATTTGGTCAAGAGAGCTCTTGGAAAAAGTAGGGGATGTTATGTTCTTTGAGGAGATGGA 1212

QY 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413
DB 1213 GTTTCCTGTTATATCCCGCCAGGATCTTCACGAGAAACAGCAGCAGGCGATTCATGAG 1272
QY 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433
DB 1273 AAATGGCCAGCAGGAAAGTTTATCTCATAGCTAGAGCTAGAGAACTGACCTCCCA 1332
QY 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453
DB 1333 TCTAGTCATCAGAAATTTTACAATAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGAT 1392
QY 454 LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473
DB 1393 TTACTTGTATAATAAGGTCGACAAAGTTGATTTTGAGGCTTTGGATAACCTGTGTGAACGA 1452
QY 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla 493
DB 1453 CTCAGGATGTCCTCAAGTGATAAGTCAAGTTAGTGATGATATTTCTTGCCCTTCTAGCT 1512
QY 494 ProLeThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513
DB 1513 CCGATTCGTATCCAGAACGTTTAGCAAAACCAATGGCGCAATTTACCTACACATGATGAT 1572
QY 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrSerAspGlyTyrIlePheAsp 533
DB 1573 GAGATTCAGTAGCCCAAGTTGGCAGCAGGATACCAACAGAGCGGTATATCTTTGAT 1632
QY 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553
DB 1633 CCTCGTATATAACCAAGTCATGATGAGGGGATGCTATGTAACTCCACATATGACCCATAGC 1692
QY 554 HisTrpIleGlyLysAspSerLeuSerLysGluLysValAlaAlaGlnAlaTyrThr 573
DB 1693 CACTCGATTAATAAAGATAGTTTGTGAAGCTGAGAGCGCGCAGCCGAGCTTATGCT 1752
QY 574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593
DB 1753 AAAGAGAAGGTTTGACCCCTCTTCGACAGCCATCAGGATTCAGAAATACTGAGGCA 1812
QY 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg 613
DB 1813 AAAGAGCAGAGAGCTATCTACAACCGCTGAAAGCAGCTAAGAGGTCGCACTTGATCGT 1872
QY 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633
DB 1873 ATGCCTTACAATCTTCAATATACTGTAGAAGTCAAAACGGTAGTTTAAATCATACCTCAT 1932
QY 634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653
DB 1933 TATGACCATACCAATCAACATCAATTTGATGTTGAGAGGCGCTTTATGAGCACT 1992
QY 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673
DB 1993 AAGGGGTATCTCTTGAGGATCTTTGGCGACTGTCAAGTACTGTGCAAGATCCAAAC 2052
QY 674 GluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLys 693
DB 2053 GAACGTCGCAATTCAGATAATGGTTTGTGTAACGCTAGGACCATGTTCAAGAAACAAA 2112
QY 694 AspHisSerGluAspProAsnLysAsnPheLysAlaAspGluGlu----- 708
DB 2113 ATGTCAGCTGATACCATCAACAGGAAACCAACCAAGCAGGAGAAACCTCAGACAGAA 2172
QY 709 ---ProValGluGluThr----- 713
DB 2173 AAACCTGAGGAAGAAACCCCTCGAGAAAGAAACCCGAAAGCGAGAAACAGAGTCTCCA 2232
QY 714 -----ProAlaGluProGluValProGlnValGluThr 724
DB 2233 AAACCAACAGAGAAACAGAGAGAAATCAGAGGAATCAGAGAACTCAGGTCGAGACT 2292

QY 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer 744
DB 2293 GAAAGGTTGAAGAAAACACTGAGAGAGCTGAAGATTTACTTGGAAAAATCCAGGATCCA 2352
QY 745 SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGln 764
DB 2353 ATTATCAAGTCCATGCAAGAGACTCTCACAGGATTAAAAAATAATTTACTATTGGC 2412
QY 765 IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLeuLysGly 784
DB 2413 ACCAGAGACAATACTATTATGCGAGAGCTGAAAAACTATTGGCTTTATTAAAGGAG 2472
QY 785 Ser 785
DB 2473 AGT 2475
RESULT 13
ID ABX06885
AC ABX06885; standard; DNA; 2457 BP.
XX
DT 27-OCT-2003 (revised)
DT 11-FEB-2003 (first entry)
XX
DE S. pneumoniae type 4 strain coding region #1173.
XX
KW Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
KW ear infection; antiinflammatory; antibacterial; immunostimulant;
KW auditory; respiratory; gene therapy; vaccine.
XX
OS Streptococcus pneumoniae; type 4 strain.
XX
PN WO200277021-A2.
XX
PD 03-OCT-2002.
XX
PF 27-MAR-2002; 2002WO-IB002163.
XX
PR 27-MAR-2001; 2001GB-00007658.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Masignani V, Tettelin H, Fraser C;
XX
DR WPI: 2003-040579/03.
DR P-PSDB; ABU01597.
XX
PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
PT ear infection.
XX
PS Claim 6; SEQ ID NO 2345; 56pp; English.
XX
CC The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC ABS56454. Also included are an antibody which binds one of the proteins,
CC treating a patient by administering the protein, DNA or antibody (in a
CC composition), a kit comprising first and second primers, which are the
CC nucleic acid cited above or fragments between nucleotides 8-100 of a
CC sequence not defined in the specification, for amplifying a target
CC sequence contained within a Streptococcus nucleic acid sequence where
CC the first primer is substantially complementary to the target sequence
CC and the second primer is substantially complementary to the complement of
CC the target sequence, and where the parts of the primers having
CC substantial complementarity define the termini of the target sequence to
CC be amplified, assay comprising contacting a test compound with the
CC protein, and determining whether the test compound binds to the protein

CC and a *Streptococcus pneumoniae* bacterium, where one or more genes
CC encoding the proteins has been rendered inactive. The proteins, nucleic
CC acid molecules, antibody and compositions are useful as medicaments for
CC treating or preventing a disease or infection due to *Streptococcus*
CC bacteria, particularly *S. pneumoniae*, such as pneumonia, sepsis, otitis
CC media or ear infection. They are also useful in developing vaccines,
CC diagnostics and antibiotics. The methods are useful for identifying
CC immunodominant proteins. The present sequence is one of the 2489
CC identified coding region from the genomic sequence. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC http://www.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to
CC standardise OS field)

SQ Sequence 2457 BP; 836 A; 497 C; 531 G; 593 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.12e-186	Length:	2457
Score:	277.50	Matches:	546
Percent Similarity:	77.30%	Conservative:	84
Best Local Similarity:	66.99%	Mismatches:	138
Query Match:	67.17%	Indels:	47
DB:	7	Gaps:	7

US-09-765-271-56 (1-796) x ABX06885 (1-2457)

QY	1	SerTyrGluLeuGlyLeuTyrGlnAla	---	ArgThrValLysGluAsnArgValSer	19
DB	61	TCCTATGAGCTTGACGCTTACCAAGCTGGTCAGGATAAGAAGAGCTCTAATCGAGTTGCT			120
QY	20	TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer	39		
DB	121	TATATAGATGGTATGACGGCTGGTCAAAAGGCAGAAAACTTGACACCCAGATGAAGTCAGT			180
QY	40	LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnClyTyrVal	59		
DB	181	AAGAGGGAGGGGATCAACGCCGGAACAAATGTTATCAAGATTACGGATCAAGGTTATGTG			240
QY	60	ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle	79		
DB	241	ACCTCTCATGGAGACCATTATCACTATTAATGGCAGAGTTCCTTATGATGCCATCATC			300
QY	80	SerGluGlnLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn	99		
DB	301	AGTGAGAGAGCTCCTCATGAAGATCCGAATTTATCAGTTGAAGGATTCAGACATTGCTCAAT			360
QY	100	GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp	119		
DB	361	GAATCAAGGGTGGTTATGTGTCATTAAAGTAAACGGTAAATACTATGTTTACCTTAAAGGAT			420
QY	120	AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis	139		
DB	421	GCAGCTCATCGGGATATATTCGCAAAAGAACAGAGATTAAACGTCAAGACGAGAACGC			480
QY	140	SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer	159		
DB	481	AGTCATAAATCATAACTCA-----AGACGAGATAATGCTGTTGTCGACCCAGAGACC			531
QY	160	GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp	179		
DB	532	CAAGGAGCTTATACACCGGATGTGGGTATATCTTCAATGATCATCTGATATCATTTAGGAC			591
QY	180	ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu	199		
DB	592	ACGGGTGATGCTTATATCGTTCCTCAGCGGACCATTAACCATTAACATTCCTTAAGAATGAG			651
QY	200	LeuSerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer	219		
DB	652	TTATCAGCTAGCGAGTTAGTGTGTCGAGAGCCCTATTGGATGGG-----			696
QY	220	AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro	239		
DB	697	-----AAGCAGGGATTCGTCTCTTCAAGTCTTCTAGTTTAAATGCA			738

QY 600 TyrAsnArgValLysGlyGluValArgIleProLeuValArgLeuProTyrMetValGlu 619
 Db 1810 TACAACCGCGTGAAGAGAGCTAAGAGTGCCTTATGATCGTATGCCTTACAACTTCAA 1869
 QY 620 HisThrValGluValLysAsnGlyAsnGluLeuIleIleProHisLysAspHisTyrHisAsn 639
 Db 1870 TATACTAGAGTCAGAACCGTAGTTTATCATCATCTATCATCATCTATCATCATTAAC 1929
 QY 640 IleLysPheAlaTrpPheAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGlu 659
 Db 1930 ATCAAAATTGAGTGTGTTTACAGAGCGCTTTATGAGGACCTTAAGGGGTATATCTTGA 1989
 QY 660 AspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsn 679
 Db 1990 GATCTTTGGCGACTGTCTAGTACTATGTGCAACATCCAAACGACGTCGCGATTCAGAT 2049
 QY 680 AspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspPro 699
 Db 2050 AATGGTTTGGTACGCTAGCGACCATGTTCAGAGAAACAAATAATGGTCAAGCTGATACC 2109
 QY 700 AsnLysAsnPheLysAlaAspGluGlu-----ProValGluGluThr 713
 Db 2110 AATCAACGGAACAAACGAGGAGGAAACCTCAGACAGAAACCTCGAGGAGAAACC 2169
 QY 713 ----- 713
 Db 2170 CCTGAGAGAGAAACCGCAAGCGAGAAACAGAGTCTCCAAACCAACAGAGGAACCA 2229
 QY 714 -----ProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGln 730
 Db 2230 GAAGAAATCACCAGAGGAATCAGAGAAACCTCAGGTCGAGACTGAAAGAGGTTGAAGAAAA 2289
 QY 731 LeuLysGluAlaGluValLeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAla 750
 Db 2290 CTGAGAGGCTGAAGATTACTTGGAAAAATCCAGGATCCAAATATCAAGTCCAAATGCC 2349
 QY 751 ThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSer 770
 Db 2350 AAAGAGACTCTCAGAGGATTAATAAATAATTTACTATTTGGCACCAGGACCAACAATACT 2409
 QY 771 IleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer 785
 Db 2410 ATTATGGCAGAGCTGAAATACTATTGGCTTTATTAAAGGAGAGT 2454
 RESULT 14
 ID AAA05417 standard; DNA; 2481 BP.
 AC AAA05417;
 DT 24-MAY-2000 (first entry)
 DE Streptococcus pneumoniae nucleotide sequence ID311.
 KW Streptococcus pneumoniae; vaccine; screening; protein antigen;
 KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;
 KW pneumococcal disease; ds.
 OS Streptococcus pneumoniae.
 PN WO200006737-A2.
 PD 10-FEB-2000.
 XX 27-JUL-1999; 99WO-GB002451.
 XX 27-JUL-1998; 98GB-00016337.
 PR 19-MAR-1999; 99US-0125164P.
 XX (MICR-) MICROBIAL TECHNIQS LTD.
 XX Gilbert CFG, Hansbro PM;
 XX

DR WPI; 2000-195300/17.
 DR P-PSDB; AAY81662.
 XX New Streptococcal protein, useful as a vaccine, for diagnosis of
 PT pneumococcal diseases and for screening agents capable of antagonizing or
 PT inhibiting expression of the protein.
 PS Claim 2; Page 99; 108pp; English.
 XX AAY81501 to AAY81679 represent specifically claimed protein sequences
 CC isolated from Streptococcus pneumoniae. AAA05407 to AAA05590 represent
 CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
 CC The sequences have antibacterial and antiinflammatory properties. The
 CC protein sequences, and fragments of them, are useful as immunogens and/or
 CC antigens. The nucleotide sequences can be used in vaccines and in
 CC diagnostic assays. The proteins and nucleotides can be useful for the
 CC detection and diagnosis of S. pneumoniae. The protein sequences are also
 CC useful for screening an agent capable of antagonizing, inhibiting or
 CC interfering with the function or expression of the proteins in which the
 CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
 CC and meningitis. AAA05591 to AAA05614 represent primers used in the
 CC exemplification of the present invention
 XX Sequence 2481 BP; 839 A; 514 C; 538 G; 590 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 1.69e-186 Length: 2481
 Score: 2795.00 Matches: 549
 Percent Similarity: 76.98% Conservative: 83
 Best Local Similarity: 66.87% Mismatches: 137
 Query Match: 67.11% Indels: 52
 DB: 3 Gaps: 8
 US-09-765-271-56 (1-796) x AAA05417 (1-2481)
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnArgValSer 19
 Db 61 TCCTATGAGCTTGGTCGTCACCAAGCTGGTCAGGATAAGAAAGATCTAATCAGTTGCT 120
 QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
 Db 121 TATATAGATGTGATCAGCTGGTCAAAAGGCGAGAAACTTGACACAGATCAAGTCAGT 180
 QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
 Db 181 AAGAGGAGGGGATCAACGCCGACAAATCGTCATCAAGATTACGATCAAGGTTATGTS 240
 QY 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79
 Db 241 ACCTCTCATGGAGACCATTTATCTATATAATGGCAAGGTCCTTATGATGCCATCATC 300
 QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
 Db 301 AGTGAAGAGCTCTCTCATGAAGATCCGAATTATCATAGTTGAAGGATTCAGACATTGTCAT 360
 QY 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
 Db 361 GAAATCAAGGGTGGTTATGTCTATCAAGGTAGACGGAATACTACTGTTTACCTTAAGGAT 420
 QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
 Db 421 GCAGCTCATGGGATAATATTCGACAAAGAGAGATTAAACGTCAGAGGAGGACGCC 480
 QY 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
 Db 481 AGTCATAATCAC-----GGTCAAGGAGCTAATCATCATCATCTAGCTAGCTCAGCAGAGCC 534
 QY 160 GlnGlyArgTyrThrThrAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
 Db 535 CAAGGACGCTATACACCGGATGATGGGTATATCTTCAATGCTCATCTATCATTTAGGAC 594
 QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199

Db 595 ACGGGTGATGCTTATATCGTTCTCCACGGCGACCATTAACCATTCCTAAGAATGAG 654
 Qy LeuSerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
 Db 655 TTATCAGCTACGAGTGTGCTGCGAAGCCTATTGGAATGGG----- 699
 Qy 220 AsnSerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239
 Db 700 -----AAGCAGGATCTCGTCTCTCTCAAGTCTTAGTTATAATGCA 741
 Qy 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
 Db 742 AATCCAGCTCAACCAAGATTCTCAGAAACACCAATCTGCTGCTCACTCCCACTTATCAT 801
 Qy 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
 Db 802 CAA---AATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATGTATGCTAAACCCCTTA 858
 Qy 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
 Db 859 TCAGACGCCATGTGGAATCTGATGGCTTATTTTCGACCAGCGCAATCACAGTCTGA 918
 Qy 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
 Db 919 ACCGCCAGAGGTGTAGCTGTCCCTCATGGTAAACCATTAACCACTTTATCCCTTTATGAACAA 978
 Qy 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
 Db 979 ATGTCTGAATGGNAACGAATTCCTGTTATTCCTTCCTTTCCTTATCGTTTCAAAACCAT 1038
 Qy 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
 Db 1039 TGGGTACCATGATTAACAGCAGAACCAACCAAGTCCCAATCGACTCCGGAACCTAGTCCA 1098
 Qy 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373
 Db 1099 AGTCGCAACCTCGACCAAAATCTCAACAGCTCCCAAGCAATCAATGTAT-----GAG 1152
 Qy 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyValPheGluGluLysGly 393
 Db 1153 AAATGGTCAAGAGCTGTTCGAAAGTAGGCGATGGTTATGTCITTTGAGGAGAAATGGA 1212
 Qy 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413
 Db 1213 GTTCTCTCGTTATATCCAGCCAAAGATCTTTCAGCAGAAACAGCAGCAGCATGTATGAGC 1272
 Qy 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433
 Db 1273 AAATCGGCCAAGCAGGAGAAAGTTTATCTCATAGCTAGGAGCTAAGAAACCTGACCTCCCA 1332
 Qy 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453
 Db 1333 TCTAGTGTATCGAATTTTACATAGGCTTATGACTTACTAGCAAGATTCACCAAGAT 1392
 Qy 454 LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473
 Db 1393 TTACTTTGATAATAAAGTCGACAAAGTTGATTTTGGAGCTTTGGATAACCTGTTTGGAAACA 1452
 Qy 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuAlaPheLeuAla 493
 Db 1453 CTAAGGATGTCCTCAAGTCAAGTAAGTCAAGTTAGTGGATGATATCTTGGCTTTCTTAGCT 1512
 Qy 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513
 Db 1513 CCGATTCGTCATCCAGACGTTTAGGNAACCAATGCGCAATTCCTACATCACTGATGAT 1572
 Qy 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533
 Db 1573 GAGATTCAAGTACCAAGTTGGCAGGCAAGTACACAAACAGACGCTTATATCTTTGAT 1632
 Qy 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553
 Db 1633 CCTCGTGATATAACCAAGTGTATGGGGGATGCGCTATGTAACTCCACATATGACCCCATAGC 1692

Qy 554 HisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr 573
 Db 1693 CACTGGATTTAAAAAAGATAGTTTCTGAAGCTGAGAGCGGAGCCAGCCCTATGCT 1752
 Qy 574 LysGluLysGlyIleLeuProSerProAspAlaAspValLysAlaAsnProThrGly 593
 Db 1753 AAAGAGAAAGTTTGACCCCTCTTCAGCAGCATCAGGATTCAGGAAATCTAGGAGCA 1812
 Qy 594 AspSerAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg 613
 Db 1813 AAAGCAGCAGAGCTATCTCAACCGCGTGAACAGCAGCTAAGAAGTGCACATTTGATCGT 1872
 Qy 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633
 Db 1873 ATGCTTTACCAATCTTCAATATCTGTAGAAGTCAAAACCGTAGTTTAAATCATACCTCAT 1932
 Qy 634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653
 Db 1933 TATGACCATTTACCATTAACATCAATTTGAGTGGTTTGACGAGGCCCTTTATGAGGCACCT 1992
 Qy 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673
 Db 1993 AAGGGGTATCTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCCAAC 2052
 Qy 674 GluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLys 693
 Db 2053 GAACGTCGCAATTCAGATAATGTTTGGTAAACGCTAGCGACCATGTGTTCAAGAAACAAA 2112
 Qy 694 AspHisSerGluAspProAsnLysAsnPheLysAlaAspGluGlu----- 708
 Db 2113 AATGGTCAAGCTGATACCATCAACCGGAAAAACCAAGCAGGAGAAACCTCAGACAGAA 2172
 Qy 709 ---ProValGluGluThr----- 713
 Db 2173 AAACCTGAGGAAGAAACCCCTCGAGAAGAGAAACCGCAAGCAGAGAACCCAGAGTCTCCA 2232
 Qy 714 -----ProAlaGluProGluValProGlnValGluThr 724
 Db 2233 AAACCAACAGAGAACCCAGAGAATCACCAGAGGAATCAGAGGAATCCTCAGTCGAGACT 2292
 Qy 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer 744
 Db 2293 GAAAGGTTGAAAGAAACTGAGAGAGGCTGAAGATTTACTTGGAAAAATCCAGGATCCA 2352
 Qy 745 SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGln 764
 Db 2353 ATTATCAAGTCCAATGCCAAAGAGACTCTCACAGGATTTAAAAAATAATTTTACTATTGGC 2412
 Qy 765 IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGly 784
 Db 2413 ACCCAGACAAACAATATCTATTATGGCAGAGCTGAAAAAATCTATTGGCTTTATTAAAGGAG 2472
 Qy 785 Ser 785
 Db 2473 AGT 2475
 RESULT 15
 AAA47605
 ID AAA47605 standard; DNA; 2531 BP.
 XX
 AC AAA47605;
 XX
 DT 20-OCT-2000 (first entry)
 XX
 DE Recombinant variant of Sp36 gene (Sp36B) of *S. pneumoniae*.
 XX
 KW Streptococcus pneumoniae; infection; vaccine; coiled coil region;
 KW histidine triad residue; Sp36; antibody; otitis media;
 KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
 XX meningitis; lobar pneumonia; ds.
 OS Streptococcus pneumoniae.

XX PH Key Location/Qualifiers
 XX FT 1. 2531
 XX FT /*tag= a
 XX FT /product= "Sp36B polypeptide"
 XX PN WO200037105-A2.
 XX XX 29-JUN-2000.
 XX XX 21-DEC-1999; 99WO-US030390.
 XX XX 21-DEC-1998; 98US-0113048P.
 XX XX (MEDI-) MEDIMMUNE INC.
 XX XX Johnson LS, Koenig S, Adamou JE;
 XX XX WPI; 2000-452129/39.
 XX XX P-PSDB; AAB01469.
 XX XX Vaccine useful for prophylaxis and treatment of pneumococcal infections
 XX PT such as otitis media, nasopharyngeal and bronchial infections, comprises
 XX PT Streptococcus pneumoniae proteins.
 XX XX Disclosure; Page 69-70; 70pp; English.
 XX XX Although a number of proteins have been suggested as being involved in
 XX CC the pathogenicity of Streptococcus pneumoniae, there still remains a need
 XX CC to identify polypeptides having epitopes in common from various strains
 XX CC of S. pneumoniae in order to utilize such polypeptides in vaccines to
 XX CC protect against a wide variety of S. pneumoniae. New vaccine compositions
 XX CC are described which comprise a Streptococcus pneumoniae polypeptide (or
 XX CC fragments) of 80 - 680 amino acids in length that comprise at least one
 XX CC histidine triad residue (HxxHxH) or a coiled-coil region, or an antibody
 XX CC directed against these features. The vaccine is useful in protecting
 XX CC against infection by Streptococcus pneumoniae. The vaccine composition
 XX CC comprising antibodies to is useful for passive immunization for treating
 XX CC pneumococcal infections which includes otitis media, nasopharyngeal and
 XX CC bronchial infections
 XX XX Sequence 2531 BP; 861 A; 508 C; 550 G; 609 T; 0 U; 3 Other;
 SQ
 Alignment Scores:
 Pred. No.: Length: 2531
 Score: 2786.50 Matches: 544
 Percent Similarity: 77.06% Conservative: 84
 Best Local Similarity: 66.75% Mismatches: 140
 Query Match: 66.90% Indels: 47
 DB: 3 Gaps: 7
 US-09-765-271-56 (1-796) x AAA47605 (1-2531)
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnAsnArgValSer 19
 Db 61 TCCTATGAGCTTGGAGCTTACCAAGCTGGTCAGGATAAGAAAGAGTCTAATCGAGTTGCT 120
 QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
 Db 121 TATATAGATGGTGATCAGCTGGTCAAAAGGGCAGAAACTTGACACCATGAAGTCAGT 180
 QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValLysIleThrAspGlnGlyTyrVal 59
 Db 181 AAGAGGGAGGGGATCAACCCGAGAACAAATGTTATCAAGATTACGATCAAGTTATGTG 240
 QY 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79
 Db 241 ACCTCTCATGGACCATTTATCATCTACTATATATGGCAAGGTTCTTATGATGCGATCATC 300
 QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
 Db 301 AGTGAAGAGCTCTCATGAAGAGATCCGAATTATCATGTTGAAGGATTCAGACATTTGCAAT 360
 QY 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
 Db 361 GAATCAAGGGTGGTATGTCATTAAGGTAACCGTAATACCTATGTTTACCTTAAGGAT 420
 QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
 Db 421 GCRGCTCATGGGATAATATTCGACAAAGAGAGATTAAACCTCGAAGACGGAACGC 480
 QY 140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
 Db 481 AGTCATAATCAATACTCA-----AGACGAGATATGCTGTGTCGACGACGAGCC 531
 QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
 Db 532 CAAGGACGTTATACAACGGATGATGGGTATATCTTCAATGTCATCTGATATCATTTGAGGAC 591
 QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
 Db 592 ACGGTGATGCTTATATGCTTCTCAGCGGACCATTTACCAATTACATTCCTTAAGAATGAG 651
 QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
 Db 652 TTATCAGCTACGAGTTAGCTGCTGCAGAGCCTATTGGAATGG----- 696
 QY 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239
 Db 697 -----AAGCAGGGATCTCTCTCTCTCAAGTTCTAGTATATAATGCA 738
 QY 240 SerValSerAsnProGlyThrThrAsnThrSerSerAsnAsnSerAsnThrAsnSer 259
 Db 739 AATCAGCTCAACCAAGATTGTCAGAGAACCAACATCTGCTGTCACCTCCACCTTATCAT 798
 QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
 Db 799 CAA---AATCAAGGGGAAACATTTCAAGCCTTTACCTGATTTGATGCTTAAACCCCTTA 855
 QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
 Db 856 TCAGAACGCCATGTGAATCTGATGGCCTTATTTTCGACCCCGCAATCACAAGTCGA 915
 QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
 Db 916 ACCGCCAGAGGTGATGCTGCTCATGTAACCAATTACCACTTTATCCCTTATGAACAA 975
 QY 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
 Db 976 ATGCTGTAATGGAATAAAGCAATTCCTCGTATTATTCCTTCCTTCGTATCGTTCAACCAT 1035
 QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
 Db 1036 TGGGTACCAAGATTCAAGACCAAGACCAAGTCCACACCAACCGACTCCAGAACCTAGTCCA 1095
 QY 360 GlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeu 379
 Db 1096 AGTCGCAACACAGCTCCCAAGCAATCCCAATTGAT-----GGGAATTTGGTCAAGAGCT 1149
 QY 380 ValArgLysValGlyGluGlyTyrValPheGluGluLysGlyLysIleSerArgTyrValPhe 399
 Db 1150 GTTCCAAAGATGAGCGATGGTTATGCTTTGAGGAGATGGAGTTTCTCGTTATATATCCCA 1209
 QY 400 AlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGlu 419
 Db 1210 GCCAAGGATCTTTCAGCAAGAACAGCAGCAGCATTTGATAGCAAACTGGCCAGCAGGAA 1269
 QY 420 SerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPhe 439
 Db 1270 AGTTTATCTCATAGCTAGAACTAGAAACCTGACCTCCCATCTAGTATGATCGAGATTT 1329
 QY 440 TyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGly 459
 Db 1330 TACAATAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACTTGTATATAAGGT 1389
 QY 460 ArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThr 479

Db	1390	CGACAAGTGTGATTTTGAGGCTTTGGATAACCTGTTGGAACGACTCAAGGATGTCTCAAGT	1449
Qy	480	AsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGlu	499
Db	1450	GATAAAGTCAAGTTAGTGGAGATATCTTCGCCCTCTTAGCTCCGATTCGTCAATCCAGAA	1509
Qy	500	ArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGln	519
Db	1510	CGTTTAGGAAACCAAAATGCGAAATACCTTACACTGATGATGAGATTCAGTAGCCCAAG	1569
Qy	520	LeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSer	539
Db	1570	TTGCAGCGAAGTACACAGCAGAGACGGTTATATCTTTGATCCTCGTGATATAACCAAGT	1629
Qy	540	AspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAsp	559
Db	1630	GATCAGGGGATGCTATGTAACTCCACATATGACCCATAGCCCTCGATTAATAAAGAT	1689
Qy	560	SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeu	579
Db	1690	AGTTTGTCTGAAGCTGAGAGCGCAGCCAGGCTTATGCTRAAGAGAAAGGTTTGACC	1749
Qy	580	ProProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaAlaIle	599
Db	1750	CTTCCTTCGACAGCACTCAGGATTCAGGAAATACTGAGGCAAAAGCAGCAGAGCTATC	1809
Qy	600	TyrAsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGlu	619
Db	1810	TACAACCGMTGAAAGCAGCTAAGAAGTGCCACTTATCATCTATGCTTACAATCTTCAA	1869
Qy	620	HisThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsn	639
Db	1870	TATACTGTAGAAGTCAAAAACGGTAGTTTAAATCATACCTCATTATGACCATTAACATAAC	1929
Qy	640	IleLysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGlu	659
Db	1930	ATCAATTTGAGTGTGTTGACGAAGGCCTTATGAGGCACCTTAAGGGGTATATCTTTGAG	1989
Qy	660	AspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsn	679
Db	1990	GATCTTTGCGACTGTCAAGTACTATGTGCAACATCCAAACGAAAGTCCGCATTTCAGAT	2049
Qy	680	AspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysAspHisSerGluAspPro	699
Db	2050	AAATGGTTTTGGTAACGCTAGCGACCATGTGTCAAAGAAAACAAAATGGTCAAGCTGATACC	2109
Qy	700	AsnLysAsnPhelLysAlaAspGluGlu-----ProValGluGluThr	713
Db	2110	AATCAACGGAAAAACCAACGGAGGAGAAACCTCAGACAGAAAAACCTGAGGAAGAACCC	2169
Qy	713	-----	713
Db	2170	CCTCGAAGAGAGAAACCGCAAGCGAGAAACACAGAGTCTCCAAAACCAACAGAGNACCA	2229
Qy	714	-----ProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGln	730
Db	2230	GAAGAAATCCACAGAGAAATCAGAAGAACCTCAGTTCGAGACTCAAAAAGGTTCAAGAAAAA	2289
Qy	731	LeuLysGluAlaGluValLeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAla	750
Db	2290	CTGAGAGAGGCTGAAGATTTACTTTGGAANAATCCAGGATCCCAATTTATCAAGTCCCAATGCC	2349
Qy	751	ThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSer	770
Db	2350	AAAGAGACTCTCACAGGATTAATAAATAATATTACTATTTTGGCACCCAGCAGCAACATACT	2409
Qy	771	IleMetAlaGluAlaGluLysLeuLeuAlaLeuLeuLysGlySer	785
Db	2410	ATTATCGCAGAGCTGAAAACTATTGGCTTTATTAAAGCAGAT	2454

RESULT 16
ABX06705

ABX06705 standard; DNA; 2517 BP.

ABX06705;

27-OCT-2003 (revised)

11-FEB-2003 (first entry)

S. pneumoniae type 4 strain coding region #933.

Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media; ear infection; antiinflammatory; antibacterial; immunostimulant; auditory; respiratory; gene therapy; vaccine.

Streptococcus pneumoniae; type 4 strain.

WO200277021-A2.

03-OCT-2002.

27-MAR-2002; 2002WO-IB002163.

27-MAR-2001; 2001GB-00007658.

(CHIR-) CHIRON SPA.

(GENO-) INST GENOMIC RES.

Masignani V, Tettelin H, Fraser C;

WPI; 2003-040579/03.

P-PSDB; AB001418.

New proteins and nucleic acid molecules from Streptococcus pneumoniae, useful as medicaments for treating or preventing a disease or infection due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or ear infection.

Claim 6; SEQ ID NO 1985; 56pp; English.

The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the specification (available on a computer readable format), or its fragment, expressed from 2469 of 2489 identified DNA coding regions from the Streptococcus pneumoniae type 4 strain genomic sequence appearing as AS56454. Also included are an antibody which binds one of the proteins, treating a patient by administering the protein, DNA or antibody (in a composition), a kit comprising first and second primers, which are the nucleic acid cited above or fragments between nucleotides 8-100 of a sequence not defined in the specification, for amplifying a target sequence contained within a Streptococcus nucleic acid sequence, where the first primer is substantially complementary to the target sequence and the second primer is substantially complementary to the target sequence and the target sequence, and where the parts of the primers having substantial complementarity define the termini of the target sequence to be amplified, assay comprising contacting a test compound with the protein, and determining whether the test compound binds to the protein and a Streptococcus pneumoniae bacterium, where one or more genes encoding the proteins has been rendered inactive. The proteins, nucleic acid molecules, antibody and compositions are useful as medicaments for treating or preventing a disease or infection due to streptococcus bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis media or ear infection. They are also useful in developing vaccines, diagnostics and antibiotics. The methods are useful for identifying immunodominant proteins. The present sequence is one of the 2489 identified coding region from the genomic sequence. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences). (Updated on 27-OCT-2003 to standardise OS field)

Alignment Scores:
Pred. No.:

SQ Sequence 2517 BP; 830 A; 509 C; 545 G; 633 T; 0 U; 0 Other;

2517

D	b		979	ATGTCCTGAATTGGAAACCAATTCGTTCGTATTTATTCCTCCTTGCCTATCGTTCAAACCAT	1038
Q	y		340	TTPValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro	359
D	b		1039	TGGGTACCAGATTCAAGACCAGAACCAACCAAGTCCACAATCGACTCCGAACTTAGTCCA	1098
Q	y		360	GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer	373
D	b		1099	AGTCCGCAACTGCACCAAAATCCTCAACGAGCTCCAAGCAATCAATTGAT-----GAG	1152
Q	y		374	SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly	393
D	b		1153	AAATTGGTCAAGAAGCTGTTTCGAAAATAGTAGCGCATGTTATGTCCTTGAGGAGAAATGGA	1212
Q	y		394	IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer	413
D	b		1213	GTITTCCTGTTAATCCACGCCCAAGATCTTTCACAGAMACGAGCAGGCAATTGATAGC	1272
Q	y		414	LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValala	433
D	b		1273	AAACTGCCCAAGCAGGAAGTTTATCTCATTAAGCTAGGAGCTAAGAAAACTGAGCTCCCA	1332
Q	y		434	ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla	453
D	b		1333	TCTAGTGATCAGAAATTTACAATAAGGCTTATGACTTACTAGCAAGNAITTCACAAGAT	1392
Q	y		454	LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg	473
D	b		1393	TTACTTGATANTAAGGTCGACAAGTTGATTTTGAGGCTTTGGATAAACCTGTGGAACGA	1452
Q	y		474	LeuAsnAspGluSerThrAsnLysGluLysLeuValAspLeuLeuAlaPheLeuAla	493
D	b		1453	CTCAAGGATGTCCTCAAGTGAATAAGTCAAGTCTAGTGGATGATATCTTTCCTCTTAGCT	1512
Q	y		494	ProIleThrHisProGluAcrLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp	513
D	b		1513	CCGATTCGTCTATCAGAACGTTTAGGAAACCAATGCGCAANATTCCTACACTGATGAT	1572
Q	y		514	GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp	533
D	b		1573	GAGATTCAGTAGGCAAGTTGGCAGGCAAGTACACACAGAAAGCGGTATATCTTTGAT	1632
Q	y		534	GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer	553
D	b		1633	CCTCGTGATATAACCAAGTGTAGGGGGATGCTATGTAACCTCCACATATGACCCCTAGC	1692
Q	y		554	HisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr	573
D	b		1693	CACTCGATTAAAAACATAGTTTTGTCTGAAGCTGAGAGCGGCGCACCGGCTTATGCT	1752
Q	y		574	LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly	593
D	b		1753	AAAGAGAAAGGTTTGACCCCTCTTCGACAGACCATCAGGATTCAGGAAATACTGAGGCA	1812
Q	y		594	AspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg	613
D	b		1813	AAAGGAGCAGAGCTATTCTACAAACCGGTGAAAGCAGCTAAGAAGGTGCCCTTATGCT	1872
Q	y		614	LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis	633
D	b		1873	ATGCTCTCAACTTCAATATACTCTAGAAGTCAAAAACGGTAGTTAATCATCATCTCAT	1932
Q	y		634	LysAspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAlaPro	653
D	b		1933	TATGACCATTCACATAACATCAAAATTGAGTCGTTTGACGAAGGCTTTATGAGGCACT	1992
Q	y		654	AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisproAsp	673
D	b		1993	AAGGGGTATACCTCTGAGGATCTTTTGGCGACTGTCAAGTACTATCTCGAACATCCAAAC	2052
Q	y		674	GluArgproHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisvalLeuGlyLys---	692

Qy	240	SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer	259
Db	742	AATCCAGCTCAACCAAGATTGTCAGAGAACCAATCTGACTGCTCACTCAACTTATCAT	801
Qy	260	GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrlsLeuProLeu	279
Db	802	CAA--AATCAAGGGGAAAAACATTCAAGCCTTTTACGTGAATGTATGTTAAACCCCTTA	858
Qy	280	SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg	299
Db	859	TCGAACGCCATCTGGAAATCTGATGGCCCTTATTTTCACCAGCGCAATACCAAGTCGA	918
Qy	300	ThrAlaArgGlyValAlaValProHisGlyAspHisTyrlsHisPheIleProTyrlsSerGln	319
Db	919	ACCGCCAGAGGTGTAGCTGCCCTCATGGTAACCAITTACCACCTTATCCCTTATGAACAA	978
Qy	320	MetSerGlnLeuGluGluArgIleAlaArgIleIleProLeuArgTyrlsArgSerAsnHis	339
Db	979	ATGTCTCAAITGGAAAAACGAATTCGTCGTATTATTCCTTCGTATCGTTCAAAACCAT	1038
Qy	340	TroValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro	359
Db	1039	TEGGTACCAGATTCAAGACCAAGAACCAACCAAGTCCCAATCGACTCCGGAACCTAGTCCA	1098
Qy	360	GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer	373
Db	1099	AGTCGGCAACCTGCACCAAAATCCTCAACGAGCTCCAAAGCAATCCAATTGAT-----GAG	1152
Qy	374	SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrlsValPheGluGluLysGly	393
Db	1153	AAATGTGTCAAAAGAGCTGTTTCGAAAAAGTAGGCGATGTTATGTTCTTTGAGGAGAATGGA	1212
Qy	394	IleSerArgTyrlsValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer	413
Db	1213	GTITCTCGTTATATCCAGCCCAAGATCTTTCAGCAGAAACAGCAGCAGCATGTATGAGC	1272
Qy	414	LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValala	433
Db	1273	AAACTGGCCCAAGCAGGAAGATTTATCTCATAAAGTAGAGCTAAGAAAACTGACCTCCCA	1332
Qy	434	ProArgAspGlnGluPheTyrlsAspLysAlaTyrlsAsnLeuLeuThrGluAlaHisLysAla	453
Db	1333	TCTAGTGTATCGAGAATTTTCAAAATAGGCTTATGACTTACTAGCAAGAAATTCACCAAGAT	1392
Qy	454	LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg	473
Db	1393	TTACTTGTATAATAAGGTCGACAAAGTTGATTTTGAGGCTTTGGATAACCTGTTGGAAACGA	1452
Qy	474	LeuAsnAspGluSerThrAsnLysGluLysLeuValAspLeuLeuAlaPheLeuAla	493
Db	1453	CTCAAGGATGTCCTCAAGTGATAAAGTCAAGTTAGTGGATGATATCTTGCTCTCTTACGT	1512
Qy	494	ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrlsThrGluAsp	513
Db	1513	CCGATTCGTATCCACAGACGTTTAGAAAAACAAATGCGCAAAATTCCTACACTGATGAT	1572
Qy	514	GluValArgIleAlaGlnLeuAlaAspLysTyrlsThrThrSerAspGlyTyrlsIlePheAsp	533
Db	1573	GAGATTCAAGTAGCCAACTTGGCAGGCAAGTACACACAGAAAGACGGTTATATCTTTGAT	1632
Qy	534	GluHisAspIleIleSerAspGluGlyAspAlaTyrlsValThrProHisMetGlyHisSer	553
Db	1633	CCTCGTGATATAACCAAGTGATGAGGGGATGCCATGTAACTCCACATATGACCCATAGC	1692
Qy	554	HisTrrIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrlsThr	573
Db	1693	CACGTGATTAATAAAGATAGTTTGTCTGAAGCTGAGAGCGCGCAGCCAGGCTTATGCT	1752
Qy	574	LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly	593
Db	1753	AAAGAAAGAGGTTTTCAGCCCTCTTTCGACAGACCATCAGGATTCAGGAATATCTAGGCA	1812

RESULT 18

RESUL 16
ABS56454 08

Continuation

WP Sequence

WP	Frequency	Frag
WP	Frequency	Frag

WP
ABS5

WP ABS5

WP ABS5

WP ABS5

WP ABS5

WP ABS5

WP ABS5

WP	ABS5
----	------

WP
ABS5

WP ABS5
WP ADCE

WP ABS5
WP ADCE

WP ABS5

WP	ABS5
WP	ABS5

WP
ABS5

WP ABS3
WP ABS5

WE
WP

WP
WP
HBS
ABS5

WP
ABS5

WP ABS5

WP	ABSS56454_20	2000001	2110000		WP	ABSS56454_21	2100001	2162598	
Alignment Scores:									
Pred. No.:	6.57e-183	Length:	110000						
Score:	2772.00	Matches:	541						
Percent Similarity:	75.78%	Conservative:	91						
Best Local Similarity:	64.87%	Mismatches:	142						
Query Match:	66.55%	Indels:	60						
DB:	7	Gaps:	8						
US-09-765-271-56 (1-796) x ABSS56454_08 (1-110000)									
QY	1	SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---	LysGluAsnAsnArgValSer	19					
DB	38009	TCCTATGAACCTGGTCTGCTCAACCAAGCTGGTCAAGTTAAGAAAGAGTCTAATCGAGTTTCT		38068					
QY	20	TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer	39						
DB	38069	TATATAGATGGTCAATCAGGCTGGTCAAAAGGCGAAGAACTTGAACACAGATGAAGTCAGT		38128					
QY	40	LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal	59						
DB	38129	AAGAGGAGGGGATCAACGCGCAACAATCGTCATCAAGATTACGATCAAGGTTATG		38188					
QY	60	ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle	79						
DB	38189	ACCTCTCATGGAGACCATATCATATTATTAATGGCAAGGTCCCTTATGATGCCATCATC		38248					
QY	80	SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn	99						
DB	38249	AGTGAGAGCTCTCATGAAGATCCGAATTATCAGTTGAAGGATTACAGACATGTGCAAT		38308					
QY	100	GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp	119						
DB	38309	GAATCAAGGGTGGTTATGTTATCAAGTAGAGTGAATACTATGTTACCTTAAGGAT		38368					
QY	120	AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis	139						
DB	38369	GCAGCTCATCGGATATATTCGGCAAAAGAGAGATTAAACGTCAAGACGAGGAACAC		38428					
QY	140	SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer	159						
DB	38429	AGTCATATCACGGGGTGGTTCT-----AACGATCAAGCAGTAGTTCCAGCCAGAGC		38482					
QY	160	GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp	179						
DB	38483	CAAGGACGCTATACACGGATGATGGTTATATCTTCAATGCTATCATATCATTTGAGGAC		38542					
QY	180	ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu	199						
DB	38543	ACGGGGTATGCTTATATCGTCTCTCAACGGCGACCATTACCATTTACATTCCTAAGATGAG		38602					
QY	200	LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer	219						
DB	38603	TTATCAGCTAGCAGTACTGCTCGAGAAGCCCTATTGGAATGGG-----	38647						
QY	220	AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro	239						
DB	38648	-----AAGCAGGGATCTCGTCCTTCTTCAAGTTCAGTTTATATGCA	38689						
QY	240	SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer	259						
DB	38690	AATCCAGCTCAACCAAGATTGTCAGAGAACCACATCTGACTGTCTCACTCCAACCTTATCAT		38749					
QY	260	GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu	279						
DB	38750	CAA---AATCAAGGGGAAACATTTCAACCCCTTTTACGTGAATTTGATGCTAAACCCCTTA		38806					
QY	280	SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg	299						
DB	38807	TCAGAACGCCATGTGGAATCTGATGGCTTTATTTTCGACCCAGCGCMAATCACAGTGA		38866					

WP	ABSS56454_20	2000001	2110000
WP	ABSS56454_21	2100001	2162598
Alignment Scores:			
Pred. No.:	6.57e-183	Length:	110000
Score:	2772.00	Matches:	541
Percent Similarity:	75.78%	Conservative:	91
Best Local Similarity:	64.87%	Mismatches:	142
Query Match:	66.55%	Indels:	60
DB:	7	Gaps:	8
US-09-765-271-56 (1-796) x ABSS56454_08 (1-110000)			
QY	1	SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---	LysGluAsnAsnArgValSer 19
DB	38009	TCCTATGAACCTGGTGCTCCCAAGCTGGTCAAGTTAAGAAGAGTCTAATCGAGTTTCT	38068
QY	20	TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer	39
DB	38069	TATATAGATGGTGATCAGGTGGTCAAAGGCCAGAAACTTGCACACAGATGAAGTCAGT	38128
QY	40	LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal	59
DB	38129	AAGAGGAGGGGATCAACGCCGACAACATCGTCATCAAGATTACGATCAAGGTTTAGTG	38188
QY	60	ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIlelle	79
DB	38189	ACCTCTCATGGAGACCATTATCATTAATAATGGCAAGGTCCCTTATGATGCCATCATC	38248
QY	80	SerGluGluLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn	99
DB	38249	AGTGAGAGCTCTCATGAAGATCCGAATATACAGTTGAAGGATTACAGACATGTGCAAT	38308
QY	100	GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp	119
DB	38309	GAATCAAGGGTGTTATGTTATCAAGTAGATGGAATACTATGTTACCTTAAGGAT	38368
QY	120	AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis	139
DB	38369	GCAGCTCATCGGATATATTCGGCAAAAGAACAGATTAAACGTCAGAGCAGGAACAC	38428
QY	140	SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer	159
DB	38429	AGTCATATCACGGGGTGTTCT-----AACGATCAGCAGTAGTTCCAGCCAGAGC	38482
QY	160	GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIlelleGluAsp	179
DB	38483	CAGGACGCTATACACGGATGATGGTTATATCTTCAATGCTCATCTATCATTTGAGGAC	38542
QY	180	ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu	199
DB	38543	ACGGGGTATGCTTATATCGTCTCTCAGCGGACCATTACCATTACATTCCTAAGATGAG	38602
QY	200	LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer	219
DB	38603	TTATCAGCTAGCAGTACTGCTGCAGAGCCCTATTGGAATGGG-----	38647
QY	220	AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro	239
DB	38648	-----AAGCAGGGATCTCGTCCTTCTCAAGTTCAGTTTATATGCA	38689
QY	240	SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer	259
DB	38690	AATCCAGCTCAACCAAGATTGTCAGAGAACCAATCTGACTGTCTCACTCCAACCTTATCAT	38749
QY	260	GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu	279
DB	38750	CAA---AATCAAGGGGAAAACATTTCAAGCCCTTTTACGTGAATTTGATGCTAAACCCCTTA	38806
QY	280	SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg	299
DB	38807	TCAGAACGCCATGTGGAATCTGATGCGCTTATTTTCGACCCAGCGCMAATCACAGTGA	38866

Db 39841 AAGGGGTATCTCTGAGATCTTTTGGCGACTGTGCACTACTATGTCACATCCAAAC 40000
 Qy 674 GluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLys 692
 Db 40001 GAACGTCGCATTTCAGATAATGGTTTGGTAACGCTAGCGACCATGTTTCGTAATAAATAG 40060
 Qy 693 -----LysAspHisSerGlu----- 697
 Db 40061 GTAGACCAAGACAGTAAACCTGATGAAGATAAGGAACATGATGAAGTAGGACCAACT 40120
 Qy 697 ----- 697
 Db 40121 CACCTGTAATCTGATGAAGAAAGAGAAATCACGCTGGTTTAAATCCTTCAGCAGATAATCTT 40180
 Qy 698 ---AspProAsnLysAsnPhelYsAlaAspGluProValGluGluThrProAlaGlu 716
 Db 40181 TATAAACCAAGCACTGATACGGAAGACAGAGAGGAAGAGTGAAGATACACAGATGAG 40240
 Qy 717 ProGluValProGluValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluVal 736
 Db 40241 GCTGAAATTCCTCAAGTAGAGAAATTCGTTATTAAACGCTAGAGATAGCAGATCGCGAGGCC 40300
 Qy 737 LeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGly 756
 Db 40301 TTGCTAGAAAGATACAGATCTCTAGTATTAGACAAATGCTATGAGACATTCAGTGGT 40360
 Qy 757 LeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGlu 776
 Db 40361 CTAAAAAGTAGTCTCTCTCGAAGCAAGAGATAATAACACTATTTCAGCAGAGTAGAT 40420
 Qy 777 LysLeuLeuAlaLeuLysGlySerAsnProSerVal 790
 Db 40421 AGTCTCTGGCTTTGTTAAAGAAAGTCAACCGGCTCTTATA 40462

RESULT 19
 AAA65737
 ID AAA65737 standard; DNA; 2639 BP.
 XX
 AC AAA65737;
 DT 21-NOV-2000 (first entry)
 XX
 DE Streptococcus pneumoniae BVH-11-2 gene SEQ ID NO:13.
 XX
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal; ds.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA001218.
 XX
 PR 23-DEC-1998; 98US-0113800P.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX
 DR WPI; 2000-452397/39.
 XX
 DR P-PSDB; AAB12720.
 XX
 XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otis media, bacteraemia and/or pneumonia.
 XX
 XX Example 6; Fig 16; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding proteins
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have

CC bactericidal activity. The nucleic acids, encoding the protein antigens,
 CC may be used for the recombinant production of the proteins they encode.
 CC The protein antigens may then be used as vaccines for the prevention and
 CC treatment of Streptococcal infections in mammals (especially humans)
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or
 CC pneumonia. The present sequence encodes the S. pneumoniae BVH-11-2
 CC protein antigen
 XX
 SQ Sequence 2639 BP; 889 A; 518 C; 567 G; 665 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 2,4e-183 Length: 2639
 Score: 2750.50 Matches: 538
 Percent Similarity: 75.30% Conservative: 90
 Best Local Similarity: 64.51% Mismatches: 145
 Query Match: 66.04% Indels: 61
 DB: 3 Gaps: 8
 US-09-765-271-56 (1-796) x AAA65737 (1-2639)
 Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19
 Db 174 TCCTATGAACTTGGTCGTCCACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTCT 233
 Qy 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
 Db 234 TATATAGATGCTGATCAGGCTGGTCAAAAGGCAGAAAATTTGACACCAGATGAAGTCAGT 293
 Qy 40 LysArgGluGlyIleAsnAlaGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
 Db 294 AAGAGAGAGGGGATCAACCGCAACAAATTTATCAAGATTACCGATCAAGTTATGTG 353
 Qy 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79
 Db 354 ACCTCTCATGAGACCATTAATCATTAATATGCAAGGTTCTTATGATGCCATCATC 413
 Qy 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
 Db 414 AGTGAAGAACTTCTCATGAAGATCCGAATTTATCATGTTGAAGGATTCAGACATTTGCAAT 473
 Qy 100 GluValLysGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
 Db 474 GAAATCAAGGGTGGCTATGATTAGGTAGCGAAGAAATATCTATGTTTACCTTAAGAT 533
 Qy 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
 Db 534 GCGGCCCATCGCACAATATTCGACAAAGAGAGATTAAACGTCAGAAGCAGGAACAC 593
 Qy 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
 Db 594 AGTCATAATCATAACTCA-----AGAGCAGATAATGCTGTGCTCAGCCAGAGCC 644
 Qy 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
 Db 645 CAAGGACGTTATACACCGGATGATGGTATATCTTCAATGCACTCATCATCATGAGGAC 704
 Qy 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
 Db 705 ACGGGTGATGCTTATATCTCTCTCACGCGCAGACCATTAACATTACATTCTTAAGAATGAG 764
 Qy 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
 Db 765 TTATCAGCTAGCGAGTTAGCTGCTGCAAGAGCTATTGGAATGGG----- 809
 Qy 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239
 Db 810 -----AAGCAGGATCTCGCTCTTCTTCAAGTTCTAGTTATATGCA 851
 Qy 240 SerValSerAsnProGlyThrThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
 Db 852 AATCCAGTTTCAACCAAGATTGTACAGAAACCAACATCTGACTGTCTCACTCAACTTATCAT 911
 Qy 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279

```

Db 912 CAA---AATCAAGGGGAAACATTTCAGCGCTTTTACGTGAATGTATGCTAAACCCCTTA 968
QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
Db 969 TCAGAACGCCATGTAGATCTGTATGATGCTTATTTTCGACCCAGCGCAATCACAAGTCGA 1028
QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
Db 1029 ACCGCCAGAGGTGTAGCTGCTCCCTCATGGTAACCAATACCACTTTATCCCTTATGAACAA 1088
QY 320 MetSerGluLeuGluGluArgIleAlaArgIleLeuProLeuAlaGlyTyrArgSerAsnHis 339
Db 1089 ATGCTGTAATTGGAAAAAGCAATGCTGTATTTATCCCTTCGTATCGTTCAACCAT 1148
QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSer--- 358
Db 1149 TGGGTACCAAGATTCAAGACCAAGAACACCAAGTCACCAATCGATCCGGAACCTAGTCCA 1208
QY 359 -----ProGlyProGlnProAlaProLeuLeuLysIleAspSerAsnSer 373
Db 1209 AGTCTGCAACCTGCACCAATCTCAACCACTCCAGCAATCCAAATGAT-----GAG 1262
QY 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393
Db 1263 AAATTGGTCAAAAGAGCTGTTTCGAAAAGTAGCGCATGTTATGCTTTGAGGAGAAATGGA 1322
QY 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413
Db 1323 GTTTCCTCGTTATATCCCAAGCAAGGATCTTTTCAGCAGAAAACAGCAGCAGGATGATAGC 1382
QY 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysLysGluAsnValAla 433
Db 1383 AACTGCGCCAGCAGGAAAGTTTATCTCATAGCTAGAGCTAAGAAAACCTGACTCCCA 1442
QY 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453
Db 1443 TCTAGTCATCGAGAATTTTACAATAAGCTTATGACTTACTAGCAAGAAATTCACCAAGAT 1502
QY 454 LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473
Db 1503 TTACTTGATATAAAGGTCGACAAAGTTGATTTGAGGTTTGGATAACCTGTTGGAACGA 1562
QY 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspLeuLeuAlaPheLeuAla 493
Db 1563 CTCAGGATGCTCAAGTATGAATAGTCAAGTATGATGATATCTTGCTTCTTAGCT 1622
QY 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513
Db 1623 CCGATTGCTCATCCAGAACGTTTAGGAAAACCAATCGCAAAATTCCTACACTGATGAT 1682
QY 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533
Db 1683 GAGATTCAAGTAGCCAAAGTTGGCAGGCAAGTACACAAAGAACGCTTATATCTTTGAT 1742
QY 534 GluHisAspIleLeuSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553
Db 1743 CCTCGTGATATAACCACTGATAGGGGGATGCCCTATGTAATCCACATATGACCCATAGC 1802
QY 554 HisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaGlnAlaTyrThr 573
Db 1803 CACTGGATTAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGAGCCAGGCTTATGCT 1862
QY 574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593
Db 1863 AAAGAGAAAGGTTTGAACCTCTCCCTTCACAGACCCACCAAGGATTCAGGAAATCTAGGCA 1922
QY 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg 613
Db 1923 AAAGGACCAAGCTATCTAACCCGGTGAAGAGCAGCTAAGAGGTCACCTTGATCGT 1982
QY 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleLeuProHis 633

```

```

Db 1983 ATGCCTTACAATCTTCAATATACCTGTAGAGTCAAAAAACGGTAGTTTAAATCATACCTCAT 2042
QY 634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653
Db 2043 TATGACCATTAACCAATCAATTCAGTTGGTTTGACGAGGCCCTTTATGAGGACCT 2102
QY 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673
Db 2103 AAGGGGTATAGTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAC 2162
QY 674 GluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLys--- 692
Db 2163 GAACGTCCGCATTCAGATATAGTTTGGTAACCGTAGTGACCATGTTCTGTAATAAATAAG 2222
QY 693 -----LysAspHisSerGlu--- 697
Db 2223 GCAGACCAAGATAGTAAACCTGTAGTAAGATAAGGAACATGATGAAGTAGTGAGCCAACT 2282
QY 697 ----- 697
Db 2283 CACCTGAATCTGATGAAAAAGAGATCAACGCTGGTTTAAATCTCTCAGCAGATAATCTT 2342
QY 698 ---AspProAsnLysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGlu 716
Db 2343 TATAAACCAAGCACTGATACGGAAGACAGACAGAGAGAAAGCTGAAGATACCAAGATGAG 2402
QY 717 ProGluValProGlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluVal 736
Db 2403 GCTGAAATCTCTCAAGTAGAAGATTCTGTTATTAACTAAGCTAAGTAGCAGATCGGAGGCC 2462
QY 737 LeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGly 756
Db 2463 TTGCTAGAAAAGGTAACAGATCTCTAGTATTAGACAAAATGCTATGAGACATTCAGCTGGT 2522
QY 757 LeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGlu 776
Db 2523 CTAAAAAGTAGTCTTCTTCGGAACGAAAGATAATAACACTATTTCAGCAGAAAGTAGAT 2582
QY 777 LysLeuLeuAlaLeuLeuLysGlySerAsnProSerSerVal 790
Db 2583 AGTCTCTGGCTTTGTTTAAAAAGAAAGTCAACCGGCTCCTATA 2624

RESULT 20
ABK15104
ID ABK15104 standard; DNA; 2639 BP.
XX
AC ABK15104;
XX
DT 08-MAY-2002 (first entry)
XX
DE DNA encoding Streptococcus pneumoniae BVH-11-2.
XX
BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;
KW streptococcal bacterial infection; gene; ds; BVH-11-2.
XX
OS Streptococcus pneumoniae.
XX
PH Key Location/Qualifiers
CDS 114..2630
FT /*tag= a
FT /product= "BVH-11-2"
FT /note= "The gene is flanked by sequences from the vector
FT S64, no information on which is given in the
FT specification"
XX
PN W0200198334-A2.
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA000098.
XX
PR 20-JUN-2000; 2000US-0212683P.
XX

```


PA (SHIR-) SHIRE BIOCHEM INC.

XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

PI WPI; 2002-122272/16.

XX P-PSDB; AAU75934.

DR

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing

PT polypeptides, useful as vaccine components for treating or preventing

PT streptococcal infections such as otitis media, meningitis, and

PT bacteremia.

XX

XX Example 3; Fig 5; 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90% identity

CC to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or

CC BVH-11, or chimeric sequences derived from them. A vaccine (II)

CC comprising (I) is useful for therapeutic or prophylactic treatment of

CC meningitis, otitis media, bacteraemia or pneumonia infection in an

CC individual susceptible to these disorders. (II) is also useful for

CC therapeutic or prophylactic treatment of any streptococcal bacterial

CC infection (e.g., caused by Streptococcus pneumoniae, group A

CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such

CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or

CC Staphylococcus aureus) in an individual susceptible to the infection. A

CC polynucleotide (III) encoding (I) is useful in DNA immunisation

CC techniques. The Streptococcus polypeptides are useful in a diagnostic

CC test for S. pneumoniae infection. (III) is useful for designing DNA

CC probes for use in detecting the presence of Streptococcus in a biological

CC sample suspected of containing the bacteria. The DNA probes may also be

CC used for detecting circulating S. pneumonia nucleic acid in a sample for

CC diagnosing streptococcal infections. This sequence encodes the

CC Streptococcus pneumonia protein BVH-11-2, used to create the antigenic

CC peptides described in the method of the invention

XX

SQ Sequence 2639 BP; 889 A; 518 C; 567 G; 665 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,4e-183 Length: 2639

Score: 2750.50 Matches: 538

Percent Similarity: 75.30% Conservative: 90

Best Local Similarity: 64.51% Mismatches: 145

Query Match: 66.04% Indels: 61

DB: 6 Gaps: 8

US-09-765-271-56 (1-796) x ABK15104 (1-2639)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19

DB 174 TCCTATGAACCTTGGTCGTCACCAAGCTGTCAGGTTAAGAAAGAGTCTAATCGAGTTTCT 233

QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39

DB 234 TATATAGATGGTGATCAGCTCGTCAAAAGGCGAGAAATTTGACACCATGAAAGTCAGT 293

QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59

DB 294 AAGAGAGGGGATCAACCCGAGACAAATTTGTTAATCAAGATTACGGATCAAGTTATGTG 353

QY 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79

DB 354 ACCTCTCATGGAGACCATATATCAATTAATGTCAGGTTCTTATGATGTCATCATC 413

QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99

DB 414 ACTGAAGAACTTCTCATGAAGATTCGAAATTAATCAAGTTGAAAGATTCAGACATTTGCAAT 473

QY 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119

DB 474 GAAATCAAGGTTGGCTATGTGATTAAGTAGAGCGGAAATATCTATGTTACCTTAAAGAT 533

QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139

DB 1563 CTCAGGATGCTCTCAAGTGATTAAGTCAAGTTAGTGGATGATATCTTCCCTTCTTAGCT 1622

DB 534 GCGGCCCATGCGGACAAATATTTCGACAAAAGAGAGATTAAACGTCAGAACAGAACAC 593

QY 140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159

DB 594 AGTCATTAATCAATACTCA-----AGACGATATATGCTGTGTCGACGACGAGCC 644

QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179

DB 645 CAGGACGTTTATCAACGATGATGGGTATATCTTCAATGATCTGATATCATTTGAGGAC 704

QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199

DB 705 ACGGTGATGCTTATATGCTTCTCAGCGGACCATTTACCATTCATTCCTTAAGAAATGAG 764

QY 200 LeuSerAlaSerGluLeuAlaAlaAlaPheLeuSerGlyArgGlyAsnLeuSer 219

DB 765 TTATCAGCTAGGATGATGCTGTCGAGAACCTTATGGAAATGGG----- 809

QY 220 AsnSerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239

DB 810 -----NAGCAGGATCTCTGCTCTTCTTCAAGTTCTAGTTATAATGCA 851

QY 240 SerValSerAsnProGlyThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259

DB 852 AATCCAGTTCACCAAGATTGTCAGAGAACCAATCTGACTGTCTCCTCCACTTATCAT 911

QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLysGlnLeuTyrLysLeuProLeu 279

DB 912 CAA---AATCAAGGGGAAACATTTCAAGCTTTTACGTGAATGTATGCTTAAACCTTA 968

QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299

DB 969 TCAGAACCCATGTAGAACTCTGATGCGCTTATTTTCGACCCAGCGCAATCACAAGTCA 1028

QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319

DB 1029 ACCGCCAGAGGTGATGCTGCTCCCTCATGTAACCATACCATTATCCCTTATGAAACA 1088

QY 320 MetSerGluLeuGluAlaArgIleAlaArgIleLeuProLeuArgTyrArgSerAsnHis 339

DB 1089 ATGCTCTGAATTTGGAATAAGCAATTTGCTGCTATTTATCCCTTCGTTATCGTTCAACCAT 1148

QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSer--- 358

DB 1149 TGGTGACAGATTCAAGACCAAGTCCACCAATTCGACTCCGGAACCTAGTCCA 1208

QY 359 -----ProGlyProGlnProAlaProAsnLeuLysIleAspSerAsnSer 373

DB 1209 AGTCTGCAACCTGCACCAATCTCAACAGCTCCAGCAATCCCAATTGAT-----GAG 1262

QY 374 SerLeuValSerGlnLeuValArgLysValGlyGlyTyrValPheGluGlyLysGly 393

DB 1263 AAATTTGGTCAAGAGCTGTTGAAAAGTAGGCGATGTTGTTGCTTTGAGGAGAATGGA 1322

QY 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413

DB 1323 GTTCTCTGTTATATCCAGCAAGGATCTTTCAGCAGAAACAGCAGCAGGCAATTGATAGC 1382

QY 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysLysGluAsnValAla 433

DB 1383 AAATCGGCCACAGCAAGAAAGTTTATCTCTAAGCTAGGAGCTAAGAAACTGACCTCCCA 1442

QY 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453

DB 1443 TCTAGTGTATCGAGATTTTACAAATTAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGAT 1502

QY 454 LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuArg 473

DB 1503 TTACTTGTAAATAAAGGTCGCAAGTTGATTTTGAAGTTTGTGATTAACCTGTTGAAACGA 1562

QY 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspLeuLeuAlaPheLeuAla 493

DB 1563 CTCAGGATGCTCTCAAGTGATTAAGTCAAGTTAGTGGATGATATCTTCCCTTCTTAGCT 1622

QY 494 ProfileThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluThrGluAsp 513
 DB 1623 CCGATTCTGTCATCCAGAACGTTTAGAAACAAATGCGCAATTAACCTACACTGATGAT 1682

QY 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533
 DB 1683 GAGATTCAAGTAGCCAGTTCGCCAGGCAAGTACACACAGAGACGGTTATATCTTTGAT 1742

QY 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553
 DB 1743 CCTCGTATATAACAGTATGAGGGGATCCCTATGTAACTCCACATATACCCCATAGC 1802

QY 554 HistripIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr 573
 DB 1803 CACTGGATTAAAGAAAGATAGTTTCTGGAAGCTGAGAGCGGAGCCAGCGCTATGCT 1862

QY 574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593
 DB 1863 AAAGAGAAAGGTTTGACCCCTCTTCGACAGACCACAGGATTTCAGGAATACTGAGGCA 1922

QY 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg 613
 DB 1923 AAGGAGCAGAGCTATCTCAACCGCGTGAAGCAGCTAAGAGGTGCCACTTGTATCGT 1982

QY 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633
 DB 1983 ATGCTTACATCTTCATATCTGTAGAGTCAAAACGGTAGTTTAATCATACCTCAT 2042

QY 634 LysAspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAlaPro 653
 DB 2043 TATGACCATTAACATAACATCAATTTGAGTGGTTTGACGAAGCCCTTTATGAGGCACCT 2102

QY 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673
 DB 2103 AAGGGGTATAGTCTGTAGGATCTTTTGCGCATCTCAAGTACTATGTCGAACATCCAAAC 2162

QY 674 GluArgProHisSerAsnAspGlyTyrGlyValAsnAlaSerGluHisValLeuGlyLys 692
 DB 2163 GAACGTCGCGATTCAGATAATGTTTGGTAAACGCTAGTACCATGTCGTAATAAATAAG 2222

QY 693 -----LysAspHisSerGlu----- 697
 DB 2223 GCAGACCAAGATAGTAAACCTGATGAAGATAAGGAACATGATGAAGTAAGTGAGCCAACT 2282

QY 697 ----- 697

DB 2283 CACCCTGAATCTGATGAAAAGAGAAATCACGCTGGTTTAAATCCTTCAGCAGATAATCTT 2342

QY 698 ---AspProAsnLysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGlu 716
 DB 2343 TATAAACCAAGACTGATACGGAAGACAGAGGAGAAAGCTGAAGATACACAGATGAG 2402

QY 717 ProGluValProGlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluVal 736
 DB 2403 GCTGAAATCTCTCAAGTAGAGAAATCTGTATTAAACGCTAAGATAGCAGATGCGAGGCC 2462

QY 737 LeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGly 756
 DB 2463 TTGCTGAAAAGTAACAGATCTCTAGTATTAGACAAATCTATGGAGACATTTGACTGGT 2522

QY 757 LeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGlu 776
 DB 2523 CTAAAGAGTAGTCTTCTCTCGNACCAAGATAATTAACACTATTTCCAGCAGAAAGTAGAT 2582

QY 777 LysLeuLeuAlaLeuLysGlySerAsnProSerSerVal 790
 DB 2583 AGTCTCTTGGCTTTGTTAAAGAAAGTCAACCGGCTCTCTATA 2624

RESULT 21
 AAV27356
 ID AAV27356 standard; DNA; 2290 BP.
 XX

AC AAV27356;
 XX 02-OCT-1998 (first entry)
 DT XX Streptococcus pneumoniae SP0042 nucleotide.
 DE XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
 KW detection; pneumonia; otitis media; meningitis; ss.
 KW XX Streptococcus pneumoniae.

OS XX Key Location/Qualifiers
 FH CDS 2..2290
 FT /*tag= a
 FT /product= "SP0042"
 FT /transl_except= (pos:152..154,aa:Xaa)
 FT /transl_except= (pos:1406..1408,aa:Xaa)
 FT /transl_except= (pos:1430..1432,aa:Xaa)
 FT /note= "no stop codon given; Xaa is unspecified"

XX W09818930-A2.

XX 07-MAY-1998.

XX 30-OCT-1997; 97WO-US019422.

XX 31-OCT-1996; 96US-0029960P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Kunsch CA, Choi GH, Johnson LS, Hromockyj A;

XX WPI; 1998-272224/24.

XX P-PSDB; AAW55095.

XX Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
 - or their epitope-containing fragments, useful in protective or
 therapeutic vaccines, and for diagnosis.

XX Claim 1; Page 61-62; 118pp; English.

XX The present sequence encodes a protein from Streptococcus pneumoniae. The
 CC nucleic acid sequence encoding the Streptococcus pneumoniae protein can
 CC be useful in vaccines for inducing protective antibodies against
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
 CC are used to detect Streptococcus infection (by usual hybridisation or
 CC amplification methods), also for isolating Streptococcus genes or their
 CC allelic variants. The protein can be used similarly to detect specific
 CC antibodies in standard immunoassays, especially for diagnosing or
 CC monitoring infections. Antibodies which bind the protein are used to
 CC detect corresponding antigens, to purify the protein and for passive
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000
 CC (especially 10-300) mu g/ml per dose

SQ Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 0 U; 5 Other;

Alignment Scores:

Pred. No.: 2,44e-176 Length: 2290
 Score: 2649.50 Matches: 519
 Percent Similarity: 76.51% Conservative: 77
 Best Local Similarity: 66.62% Mismatches: 130
 Query Match: 63.61% Indels: 53
 DB: 2 Gaps: 8

US-09-765-271-56 (1-796) x AAV27356 (1-2290)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValser 19
 DB 5 TCCTATGAACCTTGGTCGTCACCAAGCTGCTGAGTTAAGAAAGAGCTCAATCGAGTTTCT 64
 QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValser 39

1094	AAATTGGTCAAGAAAGCTGTTTCGAAAAGTAGGCGATGGTATGTATCTTTTCAGGAGAAATGGA	1155
394	ILeSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer	413
1154	GTITCTCGTTATATCCAGCGCAAGGATCTTTTCAGCAGAAACAGCAGCAGCATTTGATAGC	1213
414	LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla	433
1214	AAACTGGCAGCAGAAAGTTTATCTCTAAGCTAGGAGCTAAGAAAACTGACCTCCCA	1273
434	ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla	453
1274	TCTAGTGTATCGAGAAATTTACATAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGAT	1333
454	LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg	473
1334	TTACTTGTATATAAAGTCGACAAAGTTGATTTTTCAGGCTTTGGATACCTGTTGGAAACGA	1393
474	LeuAsnAspGluSerThrAsnLysGluLysLeuValAspLeuLeuAlaPheLeuAla	493
1394	CTCAAGATGTCNCAAGTGTAAAGTCAAGTTAGTTGGANGATATTCTTGCTTTCTTAGCT	1453
494	ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp	513
1454	CCGATTCGTCTCATCCAGAACGTTTTCAGGAAAAACCAATTCGCGAAATTCCTACACTGATGAT	1513
514	GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp	533
1514	GAGATTCAGATGCCAAGTTGGCAGGCAAGTACACAAACAGACGGGTATATATCTTTGAT	1573
534	GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer	553
1574	CCTCGTGATATAACCACTGATAGGGGGATGCCTATGTAACTCCACATATGACCCATAGC	1633
554	HisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaGlnAlaTyrThr	573
1634	CACCTGGATTAAAAAGATAGTCTTCTCAAGCTGACAGACGGCGCCAGGGCTTTATGCT	1693
574	LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly	593
1694	AAAGAGAAAGGTTTGACCCCTCTCTCCACAGACCATCAGGATTCAGGAAATACTGAGGCA	1753
594	AspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg	613
1754	AAAGGACGAAAGCTATCTCAACCGCGTGAAAGCAGCTAAGAAAGGTCGCCACTTGATCGT	1813
614	LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleLeuProHis	633
1814	ATGCTCTCAATCTTCAATATCTGTAGAAAGTCAAAACCGTAGTTTAACTATCATCTCAT	1873
634	LysAspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAlaPro	653
1874	TATGACCATTACCATAACATCAAAATTTGAGTGGTTTCAGCAAGAGCCCTTTATGAGGCACCT	1933
654	AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp	673
1934	AAGGGGTATCTCTTTGAGGATCTTTTCGGCAGCTCTCAAGTACTATGTCGACACATCCAAAC	1993
674	GluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLys	693
1994	GAACGTCGCAATTCAGATAATGGTTTGGTAAACGCTAGCAGCACTGTTCAAGAAACAAA	2053
694	AspHisSerGluAspProAsnLysAsnPhelysAlaAspGluGlu	708
2054	AATCGTCAAGCTGTATCCAAATCAAAACGGAAAAACCAAGCGAGGAGAAACCTTCAGACGAA	2113
709	---ProValGluGluThr	713
2114	AAACCTGAGGAGAAACCCCTCTGAGAGAGAAACCGCAACCGGAGAAACCTTCAGACGAA	2173
714	-----ProAlaGluProGluValProGlnValGluThr	724
2174	AAACCAACAGAGAAACCAAGAAATCACCAGAGATCAGAAAGAACCTTCAGGTGAGACT	2233

QY 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAsp 743
 DB 2234 GAAAGGTTGAGAAAAAAGCTGAGAGAGGCTGAAGATTTCCTGGAAAAAATCCAGAT 2290
 RESULT 22
 ABQ84824
 ID ABQ84824 standard; DNA; 2290 BP.
 XX AC ABQ84824;
 XX DT 04-SEP-2002 (first entry)
 XX DE S. pneumoniae SP042 nucleotide sequence SEQ ID NO:65.
 XX KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
 KW antibacterial; Streptococcal infection; detection; gene; ds.
 XX OS Streptococcus pneumoniae.
 XX PN US2002061545-A1.
 XX PD 23-MAY-2002.
 XX PF 22-JAN-2001; 2001US-00765272.
 XX PR 30-OCT-1997; 97US-00961083.
 XX PA (CHOI/) CHOI G H.
 PA (KUNS/) KUNSCH C A.
 PA (BARA/) BARASH S C.
 PA (DILL/) DILLON P J.
 PA (DOUG/) DOUGHERTY B.
 PA (FANN/) FANNON M R.
 PA (ROSE/) ROSEN C A.
 XX PI Choi GH, Kunsch CA, Barash SC, Dillon PU, Dougherty B, Fannon MR;
 PI Rosen CA;
 XX WPI: 2002-479261/51.
 DR P-PSDB; ABP54589.
 XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
 PT and for preventing or attenuating disease caused by Streptococcus
 PT infection.
 XX Claim 1; Page 28-29; 70pp; English.
 CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
 CC pneumoniae antigens have antibacterial activity and can be used in
 CC vaccines. The S. pneumoniae antigens can also be used to prevent or
 CC attenuate a Streptococcal infection in an animal. The polynucleotides
 CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
 CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
 CC of S. pneumoniae ORFs (open reading frames) which are used in an example
 CC from the present invention
 XX SQ Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 0 U; 5 Other;
 Alignment Scores:
 Pred. No.: 2,44e-176 Length: 2290
 Score: 2649.50 Matches: 519
 Percent Similarity: 76.51% Conservative: 77
 Best Local Similarity: 66.62% Mismatches: 130
 Query Match: 53
 DB: 6 Gaps: 8
 US-09-765-271-56 (1-796) x ABQ84824 (1-2290)
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19
 DB 5 TCCTATGAAGTTCGTCACCAAGCTGTCAGGTTAAGAAAGAGTCTAATCGAGTTTCT 64

QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
 DB 65 TATATAGTGTGATCAGGCTGGTCAAAGGCGAGAAAACCTTGACACACAGATGAAGTCAGT 124
 QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
 DB 125 AAGAGGGAGGGGATCAAGCGCGAACAATNGTATCAAGATTACGGATTCAAGGTTATGTG 184
 QY 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79
 DB 185 ACCTCTCATGGAGACCATTAATCTATATATGCGAAGGTTCTTATGATGCCATCATC 244
 QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
 DB 245 AGTGAAGAGCTCTCATGAAGATCCGAATTAATCAGTTGAAGGATTCAACATTGTCAAT 304
 QY 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
 DB 305 GAAATCAAGGTGGTTATGTCAATTAAGGTAAACGGTAAATACTATGTTACCTTAAGGAT 364
 QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
 DB 365 GCAGCTCATGGGATATAATTCGCAACAAGAGAGATTAAACGTCAGAGCAGAACGCC 424
 QY 140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
 DB 425 AGTCATAATCACTCACTCA-----AGAGCAGATAATGCTGTTGTCGACCCAGAGCC 475
 QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
 DB 476 CAAGGACGTTATACACGGATGATGGGTATATCTTCAATCATCATCTGATATCATGAGGAC 535
 QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
 DB 536 ACGGTCATGCTTATATCGTTCCTCAGCGCGACCATTAACATTACATTCTCTAAGAAATGAG 595
 QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
 DB 596 TTATCAGCTAGCAGATTAGCTGCTGCAGAGCCCTATTGGAATGGG----- 640
 QY 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239
 DB 641 -----AAGCAGGGATCTCGCTCTTCAAGTCTTAGTTATTAATGCA 682
 QY 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
 DB 683 AATCCAGCTCAACCAAGATTGTCAGAGAACCAATCTGACTGTCCTCAACTTATCAT 742
 QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
 DB 743 CAA---AATCAAGGGGAAACAACTTCAAGCCTTTTACGTGAATTTGTATGCTAAACCCCTTA 799
 QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
 DB 800 TCAGAACGCATGTGGAATCTGATGGCCCTTATTTTCGACCAGCGCAAAATCAAGTCGA 859
 QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
 DB 860 ACCGCCAGAGGTAGTGTGCTCCCTCATGTGTAACCATTAACCATTTATCCCTTATGACAA 919
 QY 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
 DB 920 ATGTCCTGAATTGGAAAAACGAATTCGCTATTATTATTCCTTCCTTCGTTATCGTTCAAAACCAT 979
 QY 340 TtpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
 DB 980 TGGGTACCAAGATTCAAGACCAAGAACCAAGTCCCAATCGACTCCGGAACCTAGTCCA 1039
 QY 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373
 DB 1040 AGTCGCGCAACCTGCACCAATCTCTCAACGAGCTCCAGCAATCCCAATTGAT-----GAG 1093

QY	374	SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly	393	Db	2174	AAACCAACAGAGGAACCAAGAAATCACAGAGGAATCAGAGGAACCTCAGGTCGAGACT	2233
Db	1094	AAATGGTCAAGAGAGCTGTCGAAAGTAGTAGCGAAGTGTATGCTTTGAGAGAAATGGA	1153				
QY	394	IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer	413	QY	725	GluLysValGluAlaGlnLeuLysGluAlaGluValLeuAlaLysValThrAsp	743
Db	1154	GTTCCTGTTATATCCAGCCCAAGGATCTTCAGCAGAGAAACAGCAGCAGGCGATGATAGC	1213	Db	2234	GAAAGGTTGAAGAAACACTGAGAGAGCTGAAGATTTTCTTGGAAAATCCAGGAT	2290
QY	414	LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla	433	RESULT 23			
Db	1214	AAACTGGCCAGCAGGAAAGTTATCTCATAGCTAGGAGCTAAGAAACACTGACCTCCA	1273	ADC45146			
QY	434	ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla	453	ID	ADC45146	standard; DNA; 2290 BP.	
Db	1274	TCTAGTATCCAGAAATTTTACATAGAGCTTATGACTTACTAGCAAGAAATTCACCAAGAT	1333	AC	ADC45146;		
QY	454	LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg	473	DT	18-DEC-2003	(first entry)	
Db	1334	TTACTTTGATATAATAAGGTCGACAAAGTCATTGAGGCTTTGGATAAAGCTGTGGAAAGA	1393	XX	S. pneumoniae DNA encoding antigen SP042.		
QY	474	LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla	493	XX	Antigen; ds; bacterial infection; vaccine; pneumonia; antibacterial.		
Db	1394	CTCAAGGATGTCNCAAGTGATAAAGTCAGAGTGTAGTGAGGATATTTCTTGCCTTCTTAGCT	1453	XX	Streptococcus pneumoniae.		
QY	494	ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp	513	XX	US6573082-B1.		
Db	1454	CGATTCGTATCCAGAACGTTTAGGAAACCAATGCGCAATTTACCTACACTGATGAT	1513	XX	03-JUN-2003.		
QY	514	GluValArgIleAlaGlnLeuAlaAspLysTyrThrSerAspGlyTyrIlePheAsp	533	XX	28-MAR-2000; 2000US-00536784.		
Db	1514	GAGATTCAAGTAGCCCAAGTTGCGGCAAGTACACACAGAGACGGTTATATCTTTGAT	1573	XX	31-OCT-1996; 96US-0029960P.		
QY	534	GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer	553	XX	30-OCT-1997; 97US-00961083.		
Db	1574	CCTCGTATATACACAGTGATGAGGGGATGCTATGTAACTCCACATATGACCCATAGC	1633	XX	(HUMA-) HUMAN GENOME SCI INC.		
QY	554	HisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr	573	XX	Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;		
Db	1634	CACTGGATTAAAGATAGTTGTTCTGAAGCTGAGAGCGCGACGCCAGCCTTATGCT	1693	XX	Rosen CA;		
QY	574	LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly	593	XX	WPI; 2003-764574/72.		
Db	1694	AAAGAGAAAGGTTTGACCCCTCTTCGACAGACCATCAGGATTCAGGAAATCTCAGGCA	1753	XX	P-PSDB; ADC45147.		
QY	594	AspSerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArg	613	XX	Novel polynucleotide encoding Streptococcus pneumoniae polypeptides		
Db	1754	AAAGAGCAGAGAGCTATCTACACCCGCGTGAAGCAGCTAAGAGGTGCCATTTGATGCT	1813	XX	by Streptococcus pneumoniae.		
QY	614	LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis	633	XX	Example 1; SEQ ID NO 65; 58pp; English.		
Db	1814	ATGCTTACAACTTCAATATACTGTAGAAGTCNAAACCGTAGTTTATCATACCTCAT	1873	XX	The invention relates to an isolated polynucleotide consisting of a		
QY	634	LysAspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAlaPro	653	XX	Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding		
Db	1874	TATGACCATTACCATCAATCAATTTGAGTGGTTTGACGAGCGCCCTTTATGAGGCACCT	1933	XX	SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae		
QY	654	AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrThrValGluHisProAsp	673	XX	nucleic acid into a vector, an isolated polynucleotide consisting of at		
Db	1934	AAGGGTATACCTCTTGAGGATCTTTGGCGACTGTCTAGTACTATGTCTGAAATCCAAAC	1993	XX	least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a		
QY	674	GluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLys	693	XX	recombinant host cell comprising the SP028 polynucleotide. The nucleic		
Db	1994	GAACGTCGCATTCAGATAATGGTTTGTGTAACGCTAGCGACCATGTTCAAGAAACAAA	2053	XX	acids are useful as DNA vaccine against Streptococcus pneumoniae		
QY	694	AspHisSerGluAspProAsnLysAsnPheLysAlaAspGluGlu	708	XX	infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae		
Db	2054	AATGGTCAAGCTGATACCAATCAACCGGAAAAACCAAGCGAGGAGAAACCTCAGACAGAA	2113	XX	antigen nucleic acids are useful as probes for use in diagnostic methods		
QY	709	---ProValGluGluThr	713	XX	for detecting S. pneumoniae gene expression. The present sequence encodes		
Db	2114	AAACCTGAGAGAGAAACCCCTCGAGAGAGAAACCGCAAGCGAGAAACAGAGTCTCCA	2173	XX	an S. pneumoniae antigenic protein.		
QY	714	-----ProAlaGluProGluValProGluValProGluValGluThr	724	XX	Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 0 U; 5 Other;		
				Alignment Scores:			
				Pred. No.:	2,44e-176	Length:	2290
				Score:	2649.50	Matches:	519
				Percent Similarity:	76.51%	Conservative:	77
				Best Local Similarity:	66.62%	Mismatches:	130
				Query Match:	63.61%	Indels:	53
				DB:	9	Gaps:	8
				US-09-765-271-56 (1-796) x ADC45146 (1-2290)			
QY	1	SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---	LysGluAsnAsnArgValSer	19			
Db	5	TCCTATGAACCTTGGTTCGTCCACCAAGCTGCTCAGGTTAAGAAAGAGTCTAATCGAGTTCT	64				

QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
DB 65 TATATAGATGGTGTATAGGCTGTCANAGGCAGAAACTTTGACCAGATGAAGTCACT 124
QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
DB 125 AAGAGGAGGGGATCAACGGCGAACAATNGTNAATCAAGATTACGGATCAAGGTTATGTG 184
QY 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIle 79
DB 185 ACCTCTCATGGAGACCAATTCATTTATATATGCGCAAGGTTCCTTATGATGCCATCATC 244
QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
DB 245 AGTGAAGAGCTCTCATGAAGATCCGAATTCAGTTGAAGGATTCAGACATTTGTCAT 304
QY 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
DB 305 GAAATCAAGGGTGGTTATGTCAATTAAGGTAAACGGTAAATACTATGTTNATACCTTAAGGAT 364
QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
DB 365 GCAGCTCATGCGGATATATTCGACAAAAGAGAGATTAAACGTCAGAGCAGGAACGC 424
QY 140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
DB 425 AGTCATAATCATTAATCA-----AGACGAGATAATGCTGTGTCGAGCCAGAGCC 475
QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
DB 476 CAAGGAGGTTTATACAACGGATGATGGGTATATCTTCAATGCATCTGATATCATTTAGGCAC 535
QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
DB 536 ACGGGTGATGCTTATATCTGTTCTTCAGCGGACCATTAACCATTAATTCCTTAAGAATGAG 595
QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
DB 596 TTATCAGCTAGCAGTTAGTGTGTCAGAGAGCCTATTGGATGG----- 640
QY 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239
DB 641 -----AAGCAGGATCTCGTCTCTTCAAGTCTAGTTATTAATGCA 682
QY 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
DB 683 AATCCAGCTCAACCAAGATTGTGAGAACCAACAATCTGACTGCTCACTCCAACTTATCAT 742
QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
DB 743 CAA---AATCAAGGGGAACATTTCAAGCCTTTTACGTGAATGTATGCTAAACCCCTTA 799
QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
DB 800 TCAGAACGCCATGTGGAATCTGATGGCCTTATTTTCAGCCAGCGCAAAATCAACAAGTCGA 859
QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
DB 860 ACCGCCAGAGGTGATGCTGCTCTCATGGTAACCATTAACCATTTATCCCTTATGAACAA 919
QY 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
DB 920 ATGCTGAATTTGAAAAACGAATTTGCTCGTATATTTCCCTTCTGTTATCGTTCAACCAT 979
QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
DB 980 TGGGTACCAATTCAGAACCAACCAACCAAGTCCCAATCGACTCCGGAACCTAGTCCA 1039
QY 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373
DB 1040 AGTCCGCAACCTGCACCAATTCCTCAACCAAGCTCCAGCAATTCATTTGAT-----GAG 1093
QY 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393

DB 1094 AATTTGTCAAAGAGCTGTTTCGAAAGTAGCGGATGTTTGTATGCTTTGAGGAGAAATGGA 1153
QY 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413
DB 1154 GTTTCCTGTTATATCCAGCCCAAGATCTTTCAGCAGAAACAGCAGCAGGCTTATGATAGC 1213
QY 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433
DB 1214 AAATCGCCCAAGCAGGAAGTTTATCTCATAGCTAGGAGCTAAGAAACTGACCTCCCA 1273
QY 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453
DB 1274 TCTAGTGATCGAGATTTTTCATATAGCTTATGACTTACTAGCAAGAAATTCACCAAGAT 1333
QY 454 LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuGluArg 473
DB 1334 TTACTTGATAATAAAGGTCGACAAAGTTGATTTTGAGGCTTTGGATAACCTGTTGGAACGA 1393
QY 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla 493
DB 1394 CTCAAGGATGTCNCAAGTGAATAAGTCAAGTTAGTGGAGATATTTCTTGCTTCTTAGCT 1453
QY 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513
DB 1454 CCGATTCGTATCCAGAACGTTTAGGAAACCAATTCGCAAAATTCCTACACTGATGAT 1513
QY 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533
DB 1514 GAGATTCAAGTAGCAAGTTGGCAGGCAAGTACACAACAGAGACGGTTATATCTTTGAT 1573
QY 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553
DB 1574 CCTCGTATATAACCAAGTGTATGAGGGGATGCTTATGTAATCCACATATGACCCATAGC 1633
QY 554 HistripleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr 573
DB 1634 CATGGATTAATAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGAGCCAGGCTTATGCT 1693
QY 574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593
DB 1694 AAAGAGAAAGTTTGACCCCTCTCTCGACAGACCATCAGGATTCAGGAAATACTGAGGCA 1753
QY 594 AspSerAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg 613
DB 1754 AAAGGAGCAGAAAGTATCTCAACCGCGTGAAGCAGCTAAGAAAGGTGCCACTTATGTCGT 1813
QY 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633
DB 1814 ATGCTTACATCTTCAATATCTGTAGAACTCAAAACCGGTAGTTTATATCATCTCAT 1873
QY 634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653
DB 1874 TATGACCATTTACCATTAACATCAAAATTTAGTGGTGTTCAGCAAGGCTTTTATGAGGCACT 1933
QY 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673
DB 1934 AAGGGGTATCTCTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATTCACAAAC 1993
QY 674 GluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLys 693
DB 1994 GAACGTCGGCATTCAGATTAATGTTTTGTTAGCTAGCGAGGAGAAACCTCAGACAGAA 2053
QY 694 AspHisSerGluAspProAsnLysAsnPheLysAlaAspGlu----- 708
DB 2054 AATGGTCAAGCTCATACCAATCAACGAAAAACCAAGAGGAGGAGAAACCTCAGACAGAA 2113
QY 709 ---ProValGluGluThr----- 713
DB 2114 AAACCTGAGGAAGAAACCCCTCGAGAAGAGAAACCGCAAGAGGAGAAACCGAGAGTCTCCA 2173
QY 714 -----ProAlaGluProGluValProGlnValGluThr 724

Db 2174 AAACACAGAGGACCAAGAAATCACCAGAGGAATCAGAGAACCTCAGGTGAGACT 2233
 QY 725 GluysValGluAlaGlnLeuLysGluAlaGluValLeuAlaLysValThrAsp 743
 Db 2234 GAAAGGTTGAGAAACTGAGAGAGCTGAAGATTTACTTGGAAATCCAGAT 2290
 RESULT 24
 AAX25394
 ID AAX25394 standard; DNA; 2163 BP.
 AC AAX25394;
 XX
 XX 19-JUL-1999 (first entry)
 XX Streptococcus pneumoniae complement C3-degrading protease DNA.
 XX Human complement C3-degrading protease; vaccine; infection; meningitis;
 KW pneumonia; xerotransplantation; transplant rejection; inflammation; ds.
 XX Streptococcus pneumoniae.
 OS WO9915675-Al.
 XX
 PN 01-APR-1999.
 XX
 XX 24-SEP-1998; 98WO-US020186.
 XX 24-SEP-1997; 97US-0059907P.
 XX (MINU) UNIV MINNESOTA.
 PA (AMCY) AMERICAN CYANAMID CO.
 XX
 XX Hostetter MK, Finkel DJ, Cheng Q, Green BA, Masi AW;
 XX WPI; 1999-254719/21.
 DR P-PSDB; AAY05753.
 XX
 XX New isolated human complement C3-degrading proteinase.
 PT Claim 54; Page 52-54; 65pp; English.
 XX
 PS This DNA sequence encodes a 79 kDa protein (see AAY05753) of
 CC Streptococcus pneumoniae serotype 4 that is capable of degrading human
 CC complement protein C3 (HCP3). It was identified in the S. pneumoniae
 CC serotype 4 genome by alignment to another novel open reading frame (see
 CC AAX25393) that codes for a 20 kDa HCP3 protease (AAY05752) of S.
 CC pneumoniae serotype 3. This suggested the open reading frame that codes
 CC for the 20 kDa protein may be part of a larger open reading frame. Amino
 CC acids 1-58 and 90-132 of the 20 kDa protein have substantial sequence
 CC identity with amino acids 170-227 and 258-300 of the 79 kDa protein.
 CC Proteins and peptides or polypeptides containing these regions, and DNA
 CC sequences encoding them (nucleotides 507-681 and 827-999 of the present
 CC sequence) are claimed. HCP3 proteases and polypeptides can be used as
 CC immune system stimulating compositions (claimed). They can produce an
 CC immune response against S. pneumoniae to immunize or treat a mammalian
 CC subject against infection or colonization (claimed). They can produce a B
 CC cell response, a T cell response, an epithelial cell response, or an
 CC endothelial cell response (claimed). The expression of the proteins on
 CC the surface of an organ of an animal used in xenotransplantation can be
 CC used to inhibit C3-mediated inflammation and rejection
 XX
 SQ Sequence 2163 BP; 737 A; 450 C; 462 G; 514 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 4,38e-161 Length: 2163
 Score: 2431.50 Matches: 476
 Percent Similarity: 75.41% Conservative: 79
 Best Local Similarity: 64.67% Mismatches: 130
 Query Match: 58.38% Indels: 51
 DB: 2 Gaps: 7
 US-09-765-271-56 (1-796) x AAX25394 (1-2163)

85 MetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGluValLysGlyGly 104
 Db 1 ATGAAGATCCGAATTATCAGTTGAAGGATTCAGACATTGTCAATGAATCAAGGGTGGT 60
 QY TyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAlaHisAlaAsp 124
 Db 61 TATGTTATCAAGGTAGATGGAATACTATCTTTACCTTAAGGATGCGAGCTATCGGAT 120
 QY 125 AsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSerGlnHisArgGlu 144
 Db 121 AATATTGGACAAAGAAGAGATTAAACGTGAGAAGCAGGAACACAGTCATTAATCACGG 180
 QY 145 GlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGlnGlyArgTyrThr 164
 Db 181 GGTGGTTCT-----AACGATCAAGCAGTAGTTGCGAGCCAGCCAAAGGACGCTATACA 234
 QY 165 ThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThrGlyAspAlaTyr 184
 Db 235 ACGGATGATGGTTATATCTTCAATGCATCTCATATCATTTAGAGGACACGGGTGATGCTTAT 294
 QY 185 IleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeuSerAlaSerGlu 204
 Db 295 ATCGTTCTCTACGGGACCATTTACATTACATCTCTAAGATGAGTTATCAGTAGCCGAG 354
 QY 205 LeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSerArgThrTyr 224
 Db 355 TTAGCTGTCGAGAAGCCTATTGGAATGGG----- 384
 QY 225 ArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSerValSerAsnPro 244
 Db 385 ---AAGCAGGATCTCGTCTCTTCAAGTTCTAGTTATATATGCAAAATCCAGCTCAACA 441
 QY 245 GlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAlaSerGlnSer 264
 Db 442 AGATTGTCAGAGAACCACAACTGCTGCTCACTCAACTTATCATCAA---AATCAAGG 498
 QY 265 AsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSerGlnArgHisVal 284
 Db 499 GAAACATTTCAAGCCTTTTACGTGAATTGTATGTAAACCTTATCAGAACCCATGTG 558
 QY 285 GluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThrAlaArgGlyVal 304
 Db 559 GAATCTGATGGCTTATTTTCAGCCAGCGCAAAATCACAAGTCGAACCCGAGAGGTGTA 618
 QY 305 AlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMetSerGlnLeuGlu 324
 Db 619 GCTGTCCTCATGTTAACCATTACCATTTATCCCTTATGAACAATGCTCGAATTGAA 678
 QY 325 GluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyrValProAspSer 344
 Db 679 AAACGAATTGCTCGTATTATTCCCTTCGTTATCGTTCAAAACCATTTGGTACAGATTCA 738
 QY 345 ArgProGluGlnProSerProGlnProThrProGluProSerProGlyProGlnProAla 364
 Db 739 AGACAGAACCAACCAAGTCCCAATCGACTCCGGAACCTTAGTCCAGTCCGCAACTGCA 798
 QY 365 ProAsnLeuLys-----IleAspSerAsnSerSerLeuValSerGln 378
 Db 799 CCAATCTCAACACGAGCTCCAGCAATCCAAITGAT-----GAGAAATTTGTCAAAGAA 852
 QY 379 LeuValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrVal 398
 Db 853 GCTGTTTCAAAAAGTAGGCGATGTTATGTCTTTGAGGAGAAATGGAGTTTCTGCTTATATC 912
 QY 399 PheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGln 418
 Db 913 CCAGCCAGAGATCTTTTCAGCAAAACAGCAGCAGGATTTAGTAGCAAACTGCGCCAGCAG 972
 QY 419 GluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGlu 438
 Db 973 GAAAGTTTATCTCATAAAGCTAGGAGCTAAGAAACCTGACCTCCCATCTAGTATCGAGAA 1032
 QY 439 PheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLys 458

Db 1033 TTTTACAATAAGGCTTATGACTTTACCAAGAAATTCACCAAGATTTTACTTGATAATAAA 1092
Qy GlyArgAsnSerAspPheGlnAlaLeuAspLysLeuGluArgLeuAsnAspGluSer 478
Db 1093 GGTGCAACAAGTTGATTTTGGAGCTTTGGATAACCTGTTGGACGACTCAAGGATGTCCTCA 1152
Qy ThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisPro 498
Db 1153 AGTGATAAGTCAAGTTAGTGATGATATCTTCCCTTCTTAGCTCCGATTCGTCATCCA 1212
Qy GluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAla 518
Db 1213 GAACGTTTAGGAAAAACCAATGCGCAATTTACCTACACTGATGATGATGATTCAGATGAC 1272
Qy GlnLeuAlaAspLysTyrThrSerAspGlyTyrIlePheAspGluHisAspIleIle 538
Db 1273 AAGTTGCGAGCAAGTACACAAACAGAGACGGTTATATCTTTGATCCTCGTGATATAACC 1332
Qy SerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLys 558
Db 1333 AGTGATGAGGGGATGCTATGTAATCTCCATATGACCATGACCTGATGATTAATAAAA 1392
Qy AspSerLeuSerAspLysGluLysValAlaGlnAlaTyrThrLysGluLysGlyIle 578
Db 1393 GATAGTTTGTGAAGCTGAGAGCGGCGAGCCAGGCTTATGCTAAAGAGAAAGGTTTG 1452
Qy LeuProProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaAla 598
Db 1453 ACCCTCTCTCGACAGCAATCAGGATTCAGAAATATCTGAGGCAAAAGGAGCAGAGCT 1512
Qy IleTyrAsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetVal 618
Db 1513 ATCTACACCGGTGAAGCAGCTAAGAGGTGCATTTGATCGTATGCTTACAACTCTT 1572
Qy GluHisThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHis 638
Db 1573 CAATATACTGTAGAAGTCAAAACCGTAGTTTATCATCATCTATTATGACCATTACCAT 1632
Qy AsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeu 658
Db 1633 AACATCAAAATTTGAGTGGTTTGACGAAGGCGCTTTATGAGGCACCTAAGGGGTATAC 1692
Qy GluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSer 678
Db 1693 GAGATCTTTGGCATCTGCAAGTACTATGCAATCATCAACGAAACGTCGCAATCA 1752
Qy AsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAsp 698
Db 1753 GATAATGGTTTGTGAACGCTAGCGACCATGTTTCAAAAGAAACAAATGTCAGCTGAT 1812
Qy ProAsnLysAsnPheLysAlaAspGluGlu-----ProValGluGlu 712
Db 1813 ACCAATCAAAACGAAACCAACGAGGAGAAACCTCAGACAGAAAAACCTGAGGAGAA 1872
Qy Thr----- 713
Db 1873 ACCCTCGAAGAGAAACCGAAGCGAGAAACCAAGAGTCTCCAAACCAACAGAGGAA 1932
Qy -----ProLaGluProGluValProGlnValGluThrGluLysValGluAla 729
Db 1933 CCAGAGAATCACCAGAGGAATCAGAGAACCTCAGGTCGAGACTGAAAGGTTGAAGAA 1992
Qy GlnLeuLysGluAlaGluValLeuAlaLysValThrAspSerSerLeuLysAlaAsn 749
Db 1993 AAATCTGAGAGGCTGAAGATTTTACTTGGAAAAATCCAGGATCCAAATTAAGTCCAAAT 2052
Qy AlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsn 769
Db 2053 GCCAAGAGACTCTCAGAGGATTAATAATAATTTACTATTGCGCACCAGGACCAACAT 2112
Qy SerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer 785
:::|||||

Db 2113 ACTATTATGCGAGAAGCTGAAGAACTATTGCTTTATTAAAGGAGACT 2160
RESULT 25
AAV52376
ID AAV52376 standard; DNA; 2359 BP.
XX AAV52376;
XX
XX 23-OCT-1998 (first entry)
XX
XX Streptococcus pneumoniae genome fragment SEQ ID NO:243.
XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
XX computer readable medium; vaccine; pharmaceutical composition; ds.
XX Streptococcus pneumoniae.
XX WO9818931-A2.
XX
XX 07-MAY-1998.
XX
XX 30-OCT-1997; 97WO-US019588.
XX
XX 31-OCT-1996; 96US-0029960P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;
XX Dougherty BA;
XX
XX WPI; 1998-272225/24.
XX
XX Computer-readable medium with recorded Streptococcus pneumoniae
XX polynucleotide sequences - useful in diagnostic kits and assays, and
XX pharmaceutical compositions and vaccines for Streptococcus pneumoniae.
XX
XX Claim 1; Page 1265-1266; 1409pp; English.
XX
XX The present invention describes a computer readable medium which has the
XX nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded
XX on it, or a representative fragment or a sequence at least 95% identical
XX to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
XX to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus
XX pneumoniae. The present invention also describes an isolated nucleic acid
XX molecule encoding a homologue of any of the fragments of the S. pneumoniae
XX genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
XX by a process comprising: (a) screening a genomic DNA library using as a
XX probe a target sequence defined by any of the sequences in SEQ ID NO:1 to
XX 391, identifying members of the library which contain sequences that
XX hybridize to the target sequence and isolating the nucleic acid molecules
XX from the members; or (b) isolating mRNA, DNA or cDNA produced from an
XX organism, amplifying nucleic acid molecules whose nucleotide sequence is
XX homologous to amplification primers derived from the fragment of the S.
XX pneumoniae genome to prime the amplification and isolating the amplified
XX sequences. The computer readable medium can be used in a computer-based
XX system for identifying fragments of the S. pneumoniae genome of
XX commercial importance, or expression modulating fragments of the S.
XX pneumoniae genome. Products from the present invention can be used in
XX diagnosis kits and assays, and pharmaceutical compositions and vaccines
XX for S. pneumoniae
XX
XX SQ Sequence 2359 BP; 786 A; 451 C; 511 G; 611 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 6,07e-115 Length: 2359
Score: 1774.00 Matches: 347
Percent Similarity: 81.22% Conservative: 51
Best Local Similarity: 70.82% Mismatches: 74
Query Match: 42.59% Indels: 18
DB: Gaps: 5
US-09-765-271-56 (1-796) x AAV52376 (1-2359)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 1, 2004, 07:38:42 ; Search time 4946 Seconds
(without alignments)
4805.964 Million cell updates/sec

Title: US-09-765-271-56
Perfect score: 4165
Sequence: 1 SYELGLYQARTVKENRVSY.....KLLALLKGNPSSVSXKEKIN 796

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool_h/US09765271/runat_30092004_113754_12290/app_query.fasta_1.967
-DB-EST -QFMT=fastcap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=90
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=25 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US09765271 @CIGN 1 1 3609 @runat_30092004_113754_12290 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_esti:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_pig:*
27: em_gss_vri:*
28: gb_gsal:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	177	4.2	180	28	AF325819 AF325819
2	155	3.7	2232	11	BC016682 Homo sapi
3	144	3.5	2754	29	AY416216 Mus muscu
4	144	3.5	3165	11	BC029682 Mus muscu
5	144	3.5	3166	11	AK088741 Mus muscu
6	144	3.5	3706	11	BC049893 Mus muscu
7	142	3.4	2581	11	AK003180 Mus muscu
8	141	3.4	2532	11	AK085512 Mus muscu
9	140.5	3.4	2740	11	AK029960 Mus muscu
10	139.5	3.3	3045	11	BC014691 Mus muscu
11	139.5	3.3	4984	11	AY325172 Rattus no
12	139	3.3	3573	29	AY404467 Pan trogl
13	138.5	3.3	1831	29	CG756440 P051-3-G1
14	138	3.3	3313	11	AK029864 Mus muscu
15	137	3.3	2582	11	AK016539 Mus muscu
16	137	3.3	2757	29	AY416214 Homo sapi
17	137	3.3	5304	11	BC048778 Mus muscu
18	135	3.2	2136	29	CG756344 P051-3-E1
19	135	3.2	4561	11	AK054014 Mus muscu
20	134.5	3.2	1643	29	CG753778 Mus muscu
21	134.5	3.2	1652	29	CG756662 P051-4-D0
22	134.5	3.2	1755	29	CG754413 P049-4-D0
23	134.5	3.2	4773	29	AY411030 Homo sapi
24	134	3.2	2041	29	CG755974 P051-2-E1
25	134	3.2	3714	29	AY404466 Homo sapi
26	132.5	3.2	1948	11	AK014892 Mus muscu
27	132.5	3.2	2791	11	AK081763 Mus muscu
28	132	3.2	3526	11	AK054172 Mus muscu
29	132	3.2	6914	29	AY418553 Mus muscu
30	131.5	3.2	1787	29	CG754239 P049-3-D1
31	130	3.1	3591	29	AY401924 Mus muscu
32	129.5	3.1	1439	11	AK077177 Mus muscu
33	129.5	3.1	1647	29	CG756464 P051-3-H0
34	129.5	3.1	2950	11	AK054096 Mus muscu
35	129	3.1	1744	29	CG755895 P051-2-D0
36	129	3.1	2934	11	AK089964 Mus muscu
37	129	3.1	4857	11	AK045314 Mus muscu
38	128.5	3.1	1966	29	CG756138 P051-3-A0
39	128.5	3.1	4358	11	BC035041 Homo sapi
40	128	3.1	2294	11	AY321329 Rattus no
41	128	3.1	2299	11	AY310142 Rattus no
42	128	3.1	4107	29	AY420059 Mus muscu
43	127	3.0	2067	29	CG756393 P051-3-F1
44	126.5	3.0	1797	29	CG756576 P051-4-B0
45	126.5	3.0	1877	11	AK077954 Mus muscu
46	126.5	3.0	1882	29	CG755934 P051-2-E0
47	126.5	3.0	5400	29	AY418797 Homo sapi
48	125.5	3.0	1072	29	CNS0174M Tetraodon
49	125	3.0	2175	11	BC016130 Mus muscu
50	124.5	3.0	2748	11	AK082361 Mus muscu
51	124.5	3.0	5438	29	AY402980 Mus muscu
52	124.5	3.0	6836	29	AY405716 Mus muscu
53	124.5	3.0	9744	29	AY405009 Homo sapi
54	124	3.0	954	11	CNS08XNP BX033041 Single re
55	124	3.0	1783	29	CG754377 P049-4-B0
56	123.5	3.0	1019	28	AZ685876 ENTH1177F
57	123.5	3.0	1962	29	CG756438 P051-3-G1
58	123.5	3.0	2595	11	AK028829 Mus muscu
59	123.5	3.0	2884	11	AK029095 Mus muscu
60	123.5	3.0	2981	11	AK004763 Mus muscu
61	123.5	3.0	2994	11	AK030223 Mus muscu
62	123	3.0	3582	11	AK031955 Mus muscu

```

63      123      3.0      3991      11      AK036731      Mus muscu
64      122.5      2.9      1071      10      BF237738      601842247
65      122.5      2.9      1505      10      BG024255      602302990
66      122.5      2.9      2577      29      AY406623      Homo sapi
67      122.5      2.9      3209      11      AK046752      Mus muscu
68      122.5      2.9      3541      11      AK030766      Mus muscu
69      122.5      2.9      4537      29      AY417627      Homo sapi
70      122      2.9      1785      29      CG750213      P044-3-H0
71      122      2.9      3198      11      BC009572      Homo sapi
72      121.5      2.9      1767      29      CG756340      P051-3-E0
73      121.5      2.9      2038      11      AK089164      Mus muscu
74      121.5      2.9      2218      12      BG169169      602320648
75      121.5      2.9      2380      11      AK013754      Mus muscu
76      121.5      2.9      2571      11      AK028247      Mus muscu
77      121.5      2.9      3110      11      AK019964      Mus muscu
78      120.5      2.9      1812      12      BI655534      603285602
79      120.5      2.9      1896      29      CG755802      P051-3-B0
80      120.5      2.9      3626      11      AK054303      Mus muscu
81      120      2.9      921      9      AL519296      AL519296
82      120      2.9      1515      10      BF206929      601870371
83      120      2.9      2413      11      AK019969      Mus muscu
84      120      2.9      2893      29      AY401552      Mus muscu
85      120      2.9      4137      29      AY420057      Homo sapi
86      119.5      2.9      714      12      BJ350580      BJ350580
87      119.5      2.9      1173      29      CNS06TPT      T3 end of
88      119.5      2.9      1599      10      BF981422      602309223
89      119.5      2.9      1733      29      CG756798      P051-4-G0
90      119.5      2.9      1823      29      CG756266      P051-3-D0

```

ALIGNMENTS

```

RESULT 1
AF325819      AF325819      180 bp      DNA      linear      GSS 08-MAR-2002
LOCUS
DEFINITION
  Leptospira interrogans serovar lai DNA Leptospira
  interrogans serovar lai genomic clone 18, genomic survey sequence.
ACCESSION
  AF325819
VERSION
  AF325819.1
KEYWORDS
  GSS.
SOURCE
  Leptospira interrogans serovar lai
  Leptospira interrogans serovar lai
  Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
  1 (bases 1 to 180)
  Hu, C. and Bao, J.
  Leptospira interrogans serovar lai DNA
  Unpublished (2002)
  Contact: Hu C
  Leptospiriosis Research Unit
  West China University of Medical Sciences
  17# Reming Nan Road, Chengdu, Sichuan 610041, P.R. China
  Email: huchanghua@263.net
  Class: unknown.

```

```

FEATURES
  source
  Location/Qualifiers
  1..180
  /organism="Leptospira interrogans serovar lai"
  /mol_type="genomic DNA"
  /db_xref="taxon:57678"
  /clone="18"
  /clone_lib="Leptospira interrogans serovar lai DNA"
  /note="Obtained through DNA subtraction using the
  nonpathogenic strain Leptospira biflexa serovar Patoc
  strain Patoc 1"

```

ORIGIN

```

Alignment Scores:
Pred. No.:      4.27e-05      Length:      180
Score:      177.00      Matches:      31
Percent Similarity:      74.55%      Conservative:      10
Best Local Similarity:      56.36%      Mismatches:      14
Query Match:      4.25%      Indels:      0
DB:      28      Gaps:      0

```

```

US-09-765-271-56 (1-796) x AF325819 (1-180)
Qy      84      LeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGluValLysGly 103
Db      3      TTAATGACGGACCCCAATTATCAGTTTAAACAGCTGATGTTGTTTAAACAGATTTAGAT 62
Qy      104      GlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAlaAlaHisAla 123
Db      63      GGCTATATATCAAGTTGATGCAAGTATTATGTTTATCTAAAGCCAGGAAGCAAGCGT 122
Qy      124      AspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGlu 138
Db      123      AAAATATATCCGTACCAAGCAAAATTCCTGAACAGGTTGAAAAG 167

```

RESULT 2

```

BC016682      2232 bp      mRNA      linear      HTC 05-NOV-2001
LOCUS
DEFINITION
  Homo sapiens, Similar to KIAA0164 gene product, clone
  IMAGE:3923680, mRNA.
ACCESSION
  BC016682
VERSION
  BC016682.1
KEYWORDS
  HTC.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 2232)
  Strausberg, R.
  Direct Submission
  Submitted (31-OCT-2001) National Institutes of Health, Mammalian
  Gene Collection (MGC), Cancer Genomics Office, National Cancer
  Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
  USA
  NIH-MGC Project URL: http://mgc.nci.nih.gov
  Contact: MGC help desk
  Email: cgapbs-remail.nih.gov
  Tissue Procurement: ATCC/DCFT/DTF
  cDNA Library Preparation: Life Technologies, Inc.
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Baylor College of Medicine Human Genome
  Sequencing Center
  Center code: BCM-HGSC
  Web site: http://www.hgsc.bcm.tmc.edu/cdna/
  Contact: villalona@bcm.tmc.edu.
  Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
  A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
  Muzny, D.M., Gibbs, R.A.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC plate: 26 Row: d Column: 7
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7661957
 This clone has the following problem: frame shifted.

FEATURES

```

  Location/Qualifiers
  1..2232
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="IMAGE:3923680"
  /tissue_type="Skin, melanotic melanoma."
  /clone_lib="NIH MGC_72"
  /lab_host="DH10B"
  /note="Vector: pCMV-SPORT6"

```

ORIGIN

```

Alignment Scores:
Pred. No.:      0.12      Length:      2232
Score:      155.00      Matches:      137
Percent Similarity:      32.67%      Conservative:      92
Best Local Similarity:      19.54%      Mismatches:      248
Query Match:      3.72%      Indels:      224
DB:      11      Gaps:      29

```



```

RESULT 3
LOCUS AY416216 2754 bp DNA linear GSS 12-DEC-2003
DEFINITION Mus musculus HCM5816 gene, VIRTUAL TRANSCRIPT, partial sequence,
GENOMIC survey sequence.
ACCESSION AY416216
VERSION AY416216.1 GI:39772176
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
location/Qualifiers
1..2754
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>2754
/locus_tag="HCM5816"
gene
ORIGIN
Alignment Scores:
Pred. No.: 1.31 Length: 2754
Score: 144.00 Matches: 137
Percent Similarity: 32.7% Conservative: 94
Best Local Similarity: 19.4% Mismatches: 240
Query Match: 3.46% Indels: 234
DB: 29 Gaps: 30
US-09-765-271-56 (1-796) x AY416216 (1-2754)
QY 127 ArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSerGlnHisArgGluGly 146
DB 46 AGATCACAGTCVAGTCTTAGATCAAGATCAAGATCACATTCATAGAAAGAGATACAGT 105
QY 147 ThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGlnGlyArgTyrThrThrAsp 166
DB 106 TCTAGGTCCTCGTCCAGACATATCGAGGTCCTGCTAGTAGAGATGATTTATTCTAGA 165
QY 167 Asp-----GlyTyrIlePhe 171
DB 166 GATTATCGTCGAGATTACAGGAATAATAGAGGAATGAGACGACCTTATGGGTAC----- 219
QY 172 AsnAlaSerAspIleIleGluAspThrGlyAspAlaTyrIleValProHisGlyAspHis 191
DB 220 -----AGAGAGAGGGGTAGAGGGTATTATCAAGAGGAGGAGGAGA 261
QY 192 TyrHis-----TyrIleProLysAsnGluLeuSerAlaSerGluLeuAlaAla 208
DB 262 TACCATCGAGGTGGCTATAGACCTGCTCGAATAGAGGCACTCTAGG----- 309
QY 209 GluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSerArgThrTyrArgArgGlnAsn 228

```

```

DB 310 -----AGTCCTAGACGAGGTGCGTCCAGGTCCAGGAGTCCAAAAGAGATGCC 357
QY 229 ---SerAspAsnThrSerArgThrAsnTrpValProSerValSerAsnProGlyThrThr 247
DB 358 GTGCTCTTCAAGATCCGAGGAGAGATCTCGCGGTCTATAGATCTCTAGGTCTCCA 417
QY 248 AsnThrAsnThrAsnSerAsnSerGlnAlaSerGlnAlaSerGlnSerAsnAspIle 267
DB 418 AGATCATCATCTCTCGCTCTTCATCCCATATAGCAAAATCTCTGCTCTCTAAAGACGA 477
QY 268 AspSerLeuLeuLysGlnLeuTyrIlyLeu-----ProLeu 279
DB 478 GGGTCTCAGGAAAAACAAACCAAGCTGAGGGGGAACCCCAAGAGAGAGATCCTTTG 537
QY 280 SerGlnArgHisValGlu-----SerAspGlyLeuValPheAspProAla----- 294
DB 538 AAAAGCAAAATCAGGAGGAACCAAGAGATCTTTTGAACATGATCCATCTGAATCTATT 597
QY 295 ---GlnIleThrSerArgThrAlaArgGlyValAlaValProHisGlyAspHisTyrHis 313
DB 598 GATGAGTTTAATAAATCTGCTACTTCTGCTGATATTGCGCTGGCCTTTCAGCTTATGAT 657
QY 314 PheIleProTyrSerGlnMetSerGluLeuGluGluArgIleAlaArgIleIleProLeu 333
DB 658 AATAGTCCCAAGGTCA----- 672
QY 334 ArgTyrArgSerAsnHisTrpValProAspSerArgProGluGlnProSerProGlnPro 353
DB 673 -----CCTCATAGTCTCTTCCACTTATGCT 696
QY 354 Thr-----ProGluProSerProGlyProGlnProAlaProAsnLeuLysIle----- 369
DB 697 ACACCCAGTCAGTTCATGCTCAGATGCCCATCTGCTAGTACAGTGCACCTCT 756
QY 370 -----AspSerAsnSerSerLeuValSerGlnLeuValValArgLysValGly 384
DB 757 GCCAAAAATACCCCTTCTCAGCATTCATCCATTCAGCAGCAGTCTCTGAGAGGTCTGA 816
QY 385 GluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAlaLysAspLeuPro 404
DB 817 TCTGCTCTGTT-----GGAAATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 870
QY 405 SerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSerValSerHisThr 424
DB 871 ATTCATCATCATCCCATCA-----CGAAGAAGCCCTCCCAAGACA 909
QY 425 LeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyrAspLysAlaTyr 444
DB 910 ATCACACCACAGAAAT-----GCTCCAGAGAGGAA----- 939
QY 445 AsnLeuLeuThrGluAlaHisLysAlaLeuPhe**AsnLysGlyArgAsnSerAspPhe 464
DB 940 -----TCTAGGGGGGGTCCCTCTTTAT 963
QY 465 GlnAlaLeuAsp-----LysLeuLeuGluArgLeuAsnAspGlu 477
DB 964 CCGTGAAGGAGATCAGAAACCAAGAGCAGGAAAGTTTAAAAAGGTTCACAGATGAA 1023
QY 478 SerThr-----AsnLysGluLysLeuValAsp 486
DB 1024 GAGTCTAGAGTATTCTGCTTGTATAGGGGTAAATATCAGGGATAAAGAG----- 1071
QY 487 AspLeuLeuAlaPheLeuAlaProIleThrHisProGluArgLeuGlyLysProAsnSer 506
DB 1072 -----GCTCCGAGGAGAAAGGTTCAGAAAGGCGGAGGAGGAGGAGGAGGAG 1113
QY 507 Gln-----IleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
DB 1114 GATTGGGATGACAGGAAGTCTAGATTACTTTAGTGATAAAGAGTCTCGCAAAACAA--- 1170
QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAsp----- 536
DB 1171 -----AAATTCATGACTCTGAAGGGGTGACACAGAGAGACAGAGGATTATAGACAG 1224

```

QY 537 -----llelleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHis 552
 Db 1225 TTTAGGAAATCAGTCTAGCAGATCAGGCGAGAGCTTTGCTACT----- 1269
 QY 553 SerHisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyr 572
 Db 1270 -----TCACTCACCGGAAT 1284
 QY 573 ThrLysGlnLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThr 592
 Db 1285 ACTGAGGAGGAAGCACCACAGTACAGTCC---AAAGTTTCATTAAAGGCAATAGAGAA 1341
 QY 593 GlyAspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuVal 612
 Db 1342 AGTGATGA-----TTTAGAGAGAGAAATAATATAAATGAAA 1380
 QY 613 ArgLeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIlePro 632
 Db 1381 GAGACTGCATACATAGTTCGAAAGCCCTAGCACTGCAAGGAC-----AAG 1425
 QY 633 HisLysAspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAla 652
 Db 1426 CACAAGGAA----- 1434
 QY 653 ProAsnGlyTyrThrLeuGluAspLeuPheAla-----ThrIleLysTyrTyrVal 669
 Db 1435 -----GAGACAAAGGCTCTGATAGATAACAGTCAAGAAAGAGTG 1476
 QY 670 GluHisProAspGluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisVal 689
 Db 1477 CAGTCACCGGAGCAG-----GTAAGTCTGAAAGCTC 1509
 QY 690 LeuGlyLysLysAspHisSerGluAspProAsnLysAsnPhelLysAlaAspGluPro 709
 Db 1510 AAAGAGCTCTTTGATTACAGTCCCTCTACACAGAGTCTGGACGACGAGAAAGTCC 1569
 QY 710 Val-----GluGluThrProAlaGluProGluValProGlnValGluThrGluLysVal 727
 Db 1570 ATCTTCAGAGGAGAGAGCCCATGAGATCAAAATGATAGCCAGTGTCTCACCGTCT 1629
 QY 728 GluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSerSerLys 747
 Db 1630 GAAGTCAAACTCAAAATGGACCTGTTCCTTT-----GAGATTCCACAGACCT 1680
 QY 748 AlaAsnAlaThrGlu 752
 Db 1681 GCCTCTTGACTAA 1695

RESULT 4
 LOCUS BC029682
 DEFINITION Mus musculus, 3165 bp mRNA linear HTC 07-AUG-2002
 clone IMAGE:5362245, mRNA.
 ACCESSION BC029682
 VERSION BC029682.1 GI:20987570
 KEYWORDS HTC.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 3165)
 Strauberg, R.
 Direct Submission
 Submitted (06-MAY-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 REMARK
 NTH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-x@mail.nih.gov
 Tissue Procurement: The Cepko Laboratory
 cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
 Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
 Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 53 Row: D Column: 23
 This clone was selected for full length sequencing because it
 passed the following selection criteria: GenomScan gene prediction
 This clone has the following problem: frame shifted.

FEATURES

source
 1..3165
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:5362245"
 /tissue_type="Eye, retina, mouse strain C57Bl/6"
 /clone_lib="NIH MGC_94"
 /lab_host="DH10B"
 /notes="vector: pCMV-SPORT6"

ORIGIN

Alignment Scores:
 Pred. No.: 1.61 Length: 3165
 Score: 144.00 Matches: 137
 Percent Similarity: 32.77% Conservative: 94
 Best Local Similarity: 19.43% Mismatches: 240
 Query Match: 3.46% Indels: 234
 DB: 11 Gaps: 30
 US-09-765-271-56 (1-796) x BC029682 (1-3165)
 QY 127 ArgThrLysGluGluLeuLeuAsnArgGlnLysGlnGluHisSerGlnHisArgGluGly 146
 Db 227 AGATCACAGCTAGTCTTAGATCAAGATCAAGATCAATTCATGAGAGAGAGATACAGT 286
 QY 147 ThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGlnGlyArgTyrThrAsp 166
 Db 287 TCTAGGTCTCGTTCAGAACATATTCGAGGTCTCGTAGTAGATCGTATTTTCTAGA 346
 QY 167 Asp-----GlyTyrIlePhe 171
 Db 347 GATTATCGTCGAGATTACAGGAATATAGAGGAATAGACACCTTATGGTAC----- 400
 QY 172 AsnAlaSerAspIleIleGluAspThrGlyAspAlaTyrIleValProHisGlyAspHis 191
 Db 401 -----AGAGGAAGGGGTAGAGGGTATTATCAAGGAGGAGGAGGAGA 442
 QY 192 TyrHis-----TyrIleProLysAsnGluLeuSerAlaSerGluLeuAlaAla 208
 Db 443 TACCATCGAGGTGGCTATAGACCTGTCTGGAATAGAGGCACTCTAGG----- 490
 QY 209 GluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSerArgThrTyrArgGlnAsn 228
 Db 491 -----AGTCTAGACGAGGTGGTCCGTCACGGTCCAGGAGTCCAAAGAGATCC 538
 QY 229 ---SerAspAsnThrSerArgThrAsnTrpValProSerValSerAsnProGlyThr 247
 Db 539 GTGCTCTCTCAAGATCCCGAGCAGATCTCGCGGTTCATATAGATCTCTAGTCTCTCCA 598
 QY 248 AsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAlaSerGlnSerAsnAspIle 267
 Db 599 AGATCATCATCATCTCGCTCTTCATCCCATATAGCAAAATCTCTCTCTCTCTCTCT 658
 QY 268 AspSerLeuLeuLysGlnLeuTyrLysLeu-----ProLeu 279
 Db 659 GGGTCTCAGGAAACAAACCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 718

Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861
 4
 The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5
 The PANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 3166)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishizaki, K., Nomura, K., Numazaki, R., Ono, M., Ono, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tonari, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
 Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
 Trust/MRC Building Addenbrookes Hospital Cambridge) whose
 assistance we gratefully acknowledge.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/
 Location/Qualifiers
 1. .3166
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="NOD"
 /db_xref="FANTOM,DB:E430024P04"
 /db_xref="MGI:2427638"
 /db_xref="taxon:10090"
 /clone="B430024P04"
 /cell_type="thymic cells"
 /tissue_type="thymus"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="2 days neonate"
 192. .2438
 /notes="unnamed protein product; hypothetical protein
 (SPTR) Q14673, evidence: rSCDS, NCBI CDS Predictor)
 putative"
 /codon_start=1
 /protein_id="BAC40542.1"
 /db_xref="GI:26353824"
 /translation="MGRSRSRSHSRSSKRSOSSSRSSRSHSRKRYRSRSTYSRS
 RSRDRYSRDRYDNNRMPYGRGGRGYQGGGRHRCGRYRVRNRRSRS
 PRGRSRSRSPKRSVSQRSSRRSRSSRSPSSSSSPSSPSKRSRGSQ
 EKQTKAEGRPEQESLQSEPKDTEFDHPSIDFNKNSGDIWPLGLSAYDN
 SRSPHSPIATPPQSSSCSDAPMLSTVPAKNTPSQSHSIQHSRSPSGSVGN
 GSRYSQSPSQNSPTHHIPSRSPAKTITPQNAPEERSGRSSSVPEQDQETAKTGNFLK

FEATURES

source
 CDS
 polyA_signal
 /note="putative"
 3145. .3150
 polyA_site
 3166
 /note="putative"
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.61 Length: 3166
 Score: 144.00 Matches: 137
 Percent Similarity: 32.91% Conservative: 97
 Best Local Similarity: 19.27% Mismatches: 243
 Query Match: 3.46% Indels: 234
 DB: 11 Gaps: 30
 US-09-765-271-56 (1-796) x AK088741 (1-3166)
 QY 121 AlaHisAlaAspAsnValArgThrLysGluGluLeuAsnArgGlnLysGlnGluHisSer 140
 Db 213 TCACATTCTTCAAGATCCAAAGTCTAGATCACAGCTTAGTTCTTAGATCAAGATCA 272
 QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
 Db 273 CATCTAGAAGAGAGATACAGGTCTGCTTCAGAACATATTCGAGGTCTCTGTAGTAG 332
 QY 161 GlyArgTyrThrThrAspAsp 167
 Db 333 GATCGTATTATTCTAGAGATTATCTCGAGATTACAGGAATAATAGAGAAATGAGACGA 392
 QY 168 -----GlyTyrIlePheAsnAlaSerAspIleIleGluAspThrGlyAspAlaTyrIle 185
 Db 393 CCTTATGGGTAC-----AGAGGAAGGGGTAGAGGGTATTAT 428
 QY 186 ValProHisGlyAspHisTyrHis-----TyrIleProLysAsnGluLeuSerAla 202
 Db 429 CAAGGAGGAGGAGGAGATACCATCGAGTGGCTATAGACCTCTCGGATAGAAAGGCAC 488
 QY 203 SerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSerArg 222
 Db 489 TCTAGG-----AGTCTAGACAGGTCTCGGTCCAGGTCCAGG 524
 QY 223 ThrTyrArgArgGlnAsn---SerAspAsnThrSerArgThrAsnTrpValProSerVal 241
 Db 525 AGTCCAAAAGAGAGATCCGTGCTCTCAAGATCCCGAAGCAGATCTCCCGGTCTATAT 584
 QY 242 SerAsnProGlyThrThrAsnThrSerAsnAsnSerAsnSerAsnSerGlnAla 261
 Db 585 AGATCTCTAGGTCTCCAAGATCAATCATCTCGCTCTTCATCCCATATACCAATCT 644
 QY 262 SerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeu----- 277
 Db 645 CCTGTCTCTAAAAGACGAGGTCTCAGGAAAACAAACCAAAAGCTGAGGGGGAAACCC 704
 QY 278 -----ProLeuSerGlnArgHisValGlu-----SerAspGlyLeuValPhe 291
 Db 705 CAAGAAGAGAGGTCTTTGAAAGAGCAATCACAGGAGGAACCAAGAGATATCTTTGAACAT 764
 QY 292 AspProAla-----GlnIleThrSerArgThrAlaArgGlyValAlaValPro 307
 Db 765 GATCCATCTGAATCTATTGATGAGTTTAATAATCTGCTACTCTCTGCTGATATTTGGCT 824
 QY 308 HisGlyAspHisTyrHisPheIleProTyrSerGlnMetSerGluLeuGluArgIle 327
 Db 825 GGCCTTTCAGCTTATGATAATAGTCCAAGTCA----- 857
 QY 328 AlaArgIleLeuProLeuArgTyrArgSerAsnHisTrpValProAspSerArgProGlu 347

```

Db      858  -----CCTCAT 863
Qy      348  GlnProSerProGlnProThr---ProGluProSerProGlyProGlnProAlaProAsn 366
Db      864  AGTCCTTACCTATTGCTATACACACCCAGTCAGATTCATCATGCTCAGATGCCCCCATG 923
Qy      367  LeuLysIle-----AspSerAsnSerSerLeuValSerGln 378
Db      924  CTTAGTACAGTGCACCTCTGCCAAATAATCCCTCTCTCAGCAATCATCATTCATTCAGCAC 983
Qy      379  LeuValAlaGlyLysValGlyGlyTyrValPheGluGluLysGlyIleSerArgTyrVal 398
Db      984  AGTCCTGAGAGGTCTGGATCTGGCTCTGTT-----GGAAATGGTGGTGCATATAGT 1037
Qy      399  PheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGln 418
Db      1038  CTTTCTCAGAATAGTCCGATTTCATCATCCATCCATCA-----CGA 1076
Qy      419  GluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGlu 438
Db      1077  AGAGCCCTGCCAGACAAATCACCCACAGAT-----GCTCCAGAGAGGAA--- 1124
Qy      439  PheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLys 458
Db      1125  -----TCTAGG 1130
Qy      459  GlyArgAsnSerAspPheGlnAlaLeuAsp-----LysLeuLeu 471
Db      1131  GGGGGTCTCTCTTTTATCTCTGAAGAGATCAGAAACAGCAACAGCAGGAAAGTTTTTA 1190
Qy      472  GluArgLeuAsnAspGluSerThr-----Asn 480
Db      1191  AAAGGTTTCACAGATGAAGAGTCTAGAGTATTCCTGTTGATAGGGTAATATCAGGGAT 1250
Qy      481  LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db      1251  AAAGAG-----GCTCCAGAGGAAAGGGTCAGAG 1280
Qy      501  LeuGlyLysProAsnSerGln-----IleGluTyrThrLysAspGlu 514
Db      1281  AAAGCAGGGCAGATGGAGATGGGATGACAGGAGTTCAGATTAATCTTAGTATATAA 1340
Qy      515  ValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGlu 534
Db      1341  GAGTCTGCAAAACAA-----AAATCCATGACTCTGAAGGGATCAGCAGAGGAG 1391
Qy      535  HisAsp-----IleIleSerAspGluGlyAspAlaTyrVal 546
Db      1392  ACAGAGGATTATAGACAGTTTAGGAAATCAGTCTCAGCAGATCAGGCAAGAGCTTTGCT 1451
Qy      547  ThrProHisMetGlyHisSerHisTrpIleGlyLysAspSerLeuSerAspLysGluLys 566
Db      1452  ACT-----TTTAGAGAGGAA 1547
Qy      567  ValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuProProSerProAspAlaAsp 586
Db      1455  ---TCATCTCACCGGAATACTGAGGAGGAGGAGACCCAGTCAAGTCC---AAAGTTTCA 1508
Qy      587  ValLysAlaAsnProThrGlyAspSerAlaAlaIleTyrAsnArgValLysGlyGlu 606
Db      1509  TTAAGGCAATAGAGAAAGTATGGA-----TTTAGAGAGGAA 1547
Qy      607  LysArgIleProLeuValArgLeuProTyrMetValGluHisThrValGluValLysAla 626
Db      1548  AAAAATTATAACTGAAGAGACTGCATACATAGTCGAAGCCCTAGACCTCAAGGAC 1607
Qy      627  GlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIleLysPheAlaTrpPheAsp 646
Db      1608  -----AAGCACAAGGAA 1619
Qy      647  AspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAla----- 663

```

```

Db      1620  -----GAAGACAAAGGCTCTGTAGATAATA 1643
Qy      664  ThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAspGlyTrpGly 683
Db      1644  ACAGTGAAGAAAGAGTGCAGTCCCGGAGCAG----- 1676
Qy      684  AsnAlaSerGluHisValLeuGlyLysAspHisSerGluAspProAsnLysAsnPhe 703
Db      1677  GTAAAGTCTGAAAGCTCAAAGAGCTCTTTGATTACAGTCCCTCTTACACAGAGTCTG 1736
Qy      704  LysAlaAspGluGluProVal-----GluGluThrProAlaGluProGluValProGln 721
Db      1737  GACGCACGAGAAAATCTCATCTTCAGAGGAGAGCCCACTGAGGATCAAAATGATAGCC 1796
Qy      722  ValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysVal 741
Db      1797  AGTGATCTCACCTGCTGAAGTCAACTCAAAATGGCACCCTGTTCTCTT----- 1847
Qy      742  ThrAspSerSerLeuLysAlaAsnAlaThrGlu 752
Db      1848  GACGATTCACACAGACCTGCTCTCTGACTAAA 1880

RESULT 6
BC049893      3706 bp      mRNA      linear      HTC 08-APR-2003
LOCUS      Mus musculus, Similar to protein tyrosine phosphatase, non-receptor
DEFINITION      type 3, clone IMAGE:5397303, mRNA.
ACCESSION      BC049893
VERSION      BC049893.1 GI:29612632
KEYWORDS      HTC.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 3706)
AUTHORS      Strausberg,R.
TITLE      Direct Submission
JOURNAL
COMMENT
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
CONTACT      MGC help desk
EMAIL      cgabbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dienrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Iaric,P., Legaspi,R.,
Maduro,Q.L., Mastello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: IRAK plate: 99 Row: p Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6755239
This clone has the following problem: frame shifted.
FEATURES
location/Qualifiers
1..3706
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"

```



```

/clone="IMAGE:5397303"
/tissue type="mammary tumor, C3(1)-Tag model. Infiltrating
ductal carcinoma 5 month old virgin mouse."
/clone_lib="NCI_CGAP_Mam6"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"

```

ORIGIN

Alignment Scores:

```

Pred. No.:      2.04      Length:      3706
Score:          144.00     Matches:      142
Percent Similarity: 31.96%   Conservative: 75
Best Local Similarity: 20.91% Mismatches: 234
Query Match:      3.46%     Indels:      229
DB:              11        Gaps:       33

```

US-09-765-271-56 (1-796) x BC049893 (1-3706)

```

QY 148 ProArgAsnAspGlyAlaValAlaLeuAlaAArgSerGlnGlyArgTyrThrThrAspAsp 167
DB 1119 CCTGCAAAAGCGAGCGCGCCGCTGAACCTCAGTCAGGAGGACTC----- 1163
QY 168 GlyTyrIlePheAsnAlaSerAspIleleGluAspThrGlyAspAlaTyrIleValPro 187
DB 1164 -----ATCCTCTAGGATGTCTCTGACAAACATCTTTGTGCC 1199
QY 188 HisGlyAspHisTyrHisTyrIleProIlyAsnGluLeuSerAlaSerGluLeuAlaA 207
DB 1200 AACAGAAATGGATTCTACTGTCACCTCCAGCAGACCGCTGGATAGAACCCAGATTGAC--- 1256
QY 208 AlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSerArgThrTyr-----Arg 225
DB 1257 -----CTCAGTGGTGGC-----ATCCGCAAGCGCAGCTGTACAGTGCACAC 1298
QY 226 ArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer-----ValSerAsn 243
DB 1299 AGTACGAATTCCTTAAACACACTTTCAGCCCTTACCTGCGAGCCCTCCCGCATGCTCCCAAC 1358
QY 244 ProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAlaSerGln 263
DB 1359 CCGAGTATCACTGGCAGTGATGTCATGAGACCCGAGTACATCCCGTCCCGCCACGACAGT 1418
QY 264 -----SerAsnAspIleAspSerLeuLeuGlyGlnLeu 274
DB 1419 GCGCTCATCCCGCGTCTTACCGCCCGACCCCTGACTATGAGACCGGTGATGAGCAGCTC 1478
QY 275 TyrIlys---LeuProLeuSerGlnArgHisValGluSer----- 286
DB 1479 AACAGAGGCATGGTGACAGCAGACCGGACAGCCACCTCGCTGAGGAACCTCAACATTGGC 1538
QY 287 -----AspGlyLeuValPheAspProAlaGlnIleThrSer 298
DB 1539 AGTTCTATCGGTACAGCGCGCGCGCTTGCTGTACAGCCAGCCGCGAGATT----- 1592
QY 299 ArgThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPhe---IleProTyr 317
DB 1593 CGGAGCACCCTCAGCTCCTCTCCCGAGTCAGCCCACTACCCCTTTAACCTGAACATC 1652
QY 318 SerGlnMetSerGlu-LeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSe 337
DB 1653 AGTTTCCACAGCAGCTCTC-----CGTATCCCTACCTGCGGAGAGCGG 1697
QY 337 rAsnHisTrpValProAspSerArgProGlu-----GlnPr 349
DB 1698 CCAGTGGTGGGCGCGTG-AGTGTGCTGAGTGTGACAAATGTGCGAGCTGAGGCTCAGGA 1756
QY 349 oSerProGlnProThr-----ProGluProSerProGln 360
DB 1757 CTACCCAGCTCCAAACATTATGAAACCCAGGTGTACCGGCCACCCCGCGCTTACCTTA 1816
QY 360 yProGlnProAla-----ProAsnLeuLys- 372
DB 1817 CCCAAGACCTGCCAAACAGCAGCCCGACACCTGTCCCGCCACCTCTACATCAGCAGCAGCA 1876

```

```

QY 372 nSerSerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGlu 392
DB 1877 TCCGATCTCATCACAGCGCGTCCATCACTCACTGTCAG-----ACCTTCCAGAGGA 1930
QY 392 sGlyIleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGln 412
DB 1931 CAGCTTGGCCCGTGGTCACTCTCTGAGAGGAGGTGAGGAGCCCTCACAGCAGCGCGCA 1990
QY 412 uSerLysLeuSerLysGlnGluSerVal----- 421
DB 1991 TGCCCACTACAGAGAGAGACAGTATTGAATCGCAGGGCTCACACATGATTCGAAGG 2050
QY 422 -----SerHisThrLeuThrAlaLysLysGluAsnValAlaProArg----- 435
DB 2051 CCTGAGGCTCAAGGAGAGAGCGTGTCTGAGCCTCAGCAGCAGCGTGGCGCCCGACCTT 2110
QY 436 -----AspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAl 450
DB 2111 CTCAGCAGGCTCCAGTCCAGCTCTCTCTGACAAATGAAGCAGGAGGGGACTGAAGA 2170
QY 450 aHisLys---AlaLeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLys 469
DB 2171 GCAGGAGGTGGCAGATACAGCCATAGAAGTCCCTTCTGATGCCACCATGCTGATTCA 2230
QY 469 sLeuLeuGluArgLeuAsnAsp-----GluSerThrAsnLysGluLysLeuValAspAs 487
DB 2231 TAGCAGTGAGGAGGACGAGGACTTGGAGAGGACAGCAGCAGAGACAGGCGCATCTCAGC 2290
QY 487 pLeu-----LeuAlaPhe 491
DB 2291 TGTGCTGAGCCCGCTTACGGCGCTTCTCTCAGGAATAAACTACCCCTGTGCTTC 2350
QY 491 eLeuAlaPheIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrTh 511
DB 2351 AGCGACTCCGATCACTCGGCTCTGCTGATATTTTGGAGCCCAAGCCCATGTC-----AC 2404
QY 511 rGluAspGluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIl 531
DB 2405 AGAGCTTGAGAGAGGCGCAAGGACATCAGCCCTGTCCAC-----CTGGT 2449
QY 531 ePheAspGluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGl 551
DB 2450 TGTGGAGACCAT-----CGGCCCCGAGAGATGGAGTCTGCTACCCCTCCATGTC 2500
QY 551 yHisSerHisTrpIleGly----- 557
DB 2501 CGAGTCAGACCTCAGGACCTCAGGAGGTACCGAGCTAGGAGGAGTCTCTCAAGAAAG 2560
QY 558 -----LysAspSerLeuSerAspLysGluLysValAla----- 568
DB 2561 GCCAGTGTACAGCTCTCTCTGGGAAAAAAGCGCTGTGGAAAGGACTTCGCGCACTAGG 2620
QY 569 -----AlaGlnAlaTyrThrLysGluLysGlyIleLeuPro----- 580
DB 2621 GGGGATGAAAAAATCGAGCAGATGCAAAAAAATTTGGTCCCTCAAGCTGGCAGCCCT 2680
QY 581 -----ProSerProAsp-----AlaAspValLysAlaAs 590
DB 2681 CAATGGGCTCTCCCTGTCCCGGCTGCCCTGCTGATGAGGAAAGAGAGTGTCCACCG 2740
QY 590 nProThrGlyAspSerAlaAlaIle----- 599
DB 2741 AGCGACAAATGACGAGAGGTGCAAGGTCTTGAGCAGCGGTAGAGCAAGGATGGTCTT 2800
QY 600 -----TyrAsnArg-----ValLysGlyGlu-LysArgIlePro- 610
DB 2801 CACCGAATGAGAGGATCCTTAAAGAGCGGTGGTTGATGGCGAGTGTCTGACTGCCCG 2860
QY 611 -----LeuValArg- 613
DB 2861 GCTCCCTGAAATGACAGAGAAATCGATTTCAGATGTCTCCCTACGATGATGGAG 2920

```

Qy 614 -----LeuProTyMetValGluHisThr----- 621
 Db 2921 AGTCAGTTGGTGCACACAGGAGAACACACACCTGGTACATCAACGGCTCACACATTAA 2980
 Qy 622 -----ValGluVallyAsnGlyAsnLeuIlelleProHisLysAspHisTyr 637
 Db 2981 GGTCTCTGTCAGCGGAATAGATGGATTATATTGCCACACAGGACCATTAAC 3033

RESULT 7
 AK003180
 LOCUS
 DEFINITION Mus musculus 18-day embryo whole body cDNA, RIKEN full-length enriched library, clone:1100001J13 product:NULP1 homolog [Mus musculus], full insert sequence.
 AK003180
 VERSION AK003180.1 GI:12833685
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253
 10349636

2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 11042159

3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861

4
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)

5
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 2581)

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission

JOURNAL

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]
 Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGCGCGCACTGAGTCTTTTATTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequences [5'-GAGAGAGAGATCCAGAGCTCAATTAATTAACCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

FEATURES

source

1..2581
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM DB:1100001J13"
 /db_xref="MG1:1894979"
 /db_xref="taxon:10090"
 /clone="1100001J13"
 /tissue_type="whole body"
 /clone_libs="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="18-day embryo"
 51..1930
 /note="NULP1 homolog [Mus musculus] (SPTR|Q9ER19, evidence: FASTI, 99%ID, 99.8%length, match=1872) putative"

misc_feature

ORIGIN

Alignment Scores:
 Pred. No.: 1.72 Length: 2581
 Score: 142.00 Matches: 175
 Percent Similarity: 30.92% Conservative: 85
 Best Local Similarity: 20.81% Mismatches: 272
 Query Match: 3.41% Indels: 310
 DB: 11 Gaps: 43

US-09-765-271-56 (1-796) x AK003180 (1-2581)

Qy 12 ValLysGluAsnAsnArgValSerTyrIleAspGlyLysGlnAlaThrGlnLysThrGlu 31
 Db 216 GTCCGTGTCAACACCGGTTTGAGCTGATAAAC-----ACGGAG 254
 Qy 32 AsnLeuThrProAspGluValSerLysArgGluGlyIleAsnAlaGluGlnIleValIle 51
 Db 255 GATCTGTGAGGATGACCTTGTG-----GTAAATGGGAGAGATCTGACTGT 299
 Qy 52 LysIleThrAspGlnGlyTyrValThrSerHisGlyAspHisTyrHisTyrTyrAsnGly 71
 Db 300 ACCCTGCCGATCTCT-----GTGTCATCAGGGAAC-----AAAGGA 335
 Qy 72 LysValProTyrAspAlaIleIleSerGluGluLeuLeuMeLysAspProAsnTyrLys 91
 Db 336 AGGGCT-----AAACATGGAACACGACGAG 359
 Qy 92 LeuLysAspGluAspIleValAsnGluValLysGlyGlyTyrValIleLysValAspGly 111
 Db 360 ACCAAGCAAGATGGA-----GGGGCCACCAAGCAGG-----392
 Qy 112 LysTyrTyrValTyrLeuLysAspAlaIleHisAlaAspAsnValArgThrLysGluGlu 131
 Db 393 -----TCTCGGAACAGCTTAATGCAAGTGCACAACTCCGA 428

QY 268 -----AspSerLeuLeuLysGln 273
 Db 529 ACAGGCTCTGAGGTATTCGGCGGGGTGGTCTCCCAAGGCGGATGGACACCGCGGTCA 588
 QY 274 LeuTyrLysLeuProLeuSer----- 280
 Db 589 CTGGACTACGTGGAGCTCTCTCCATTAGCGCCAAAGCTCCCCACAGCGCATGGCACTCTC 648
 QY 281 -----GlnArgHisValGluSerAspGlyLeuValPheAspProLagln 295
 Db 649 TCCCGCTCCACTCCCGGAGCGCCCAACCAAGCAGGAGGATCTGGAGCGGGACTTGGCCCAA 708
 QY 296 IleThrSerArg----- 299
 Db 709 CGCTCTGAGGAACAGCAAGTGGTTCAGTCCACGAGCGCAGGACCCACAGACGCC 768
 QY 300 -----ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIle 315
 Db 769 AGTGGCGATGGTCTCGAAGGGGCTGGGTGCCCTCCCTGACTGATGACCAAG----- 822
 QY 316 ProTyrSerGlnMetSer---GluLeuGluGluArgIleAlaArgIleIleProLeuArg 334
 Db 823 -----ACAGGCTCAGTGAGGAGATTGAGAAGAGTGGCAGGAGTTGAAAAGCTGCC 876
 QY 335 TyrArgSerAsnHisTrpValProAspSerArgProGluGlnProSerProGln----- 352
 Db 877 CTGGGGAGAACAGCGAGTGCCTCACCCTCCCTGCTCAACCAAGCCCAACAGTCG 936
 QY 353 -----ProThr 354
 Db 937 TCGGGGACCTACAAGTACACCATGAGGCCCTGGAGAAGGAGGTCCAGTCTCTCCGTGC 996
 QY 355 Pro-----GluPro-SerProGlyProGlnProLapProAsnLeuLysII 369
 Db 997 CCAGCTGGAGCTTGGGCTCTCCCGGGGAACTCTCCAGAAATGCACCGAGACTCCAGGA 1056
 QY 369 eAspSerAsnSerSerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPh 389
 Db 1057 GGACAGCCAT-----ATCCCCCGGCTACATCTC 1086
 QY 389 eGluGlyLysGlyIleSerArgTyrValPheAlaLysAspLeuProSerGluThrVal-- 408
 Db 1087 GCAGGAA---GCCTGTGAGCGCAGCTGGCAGAGATGGAGTCCCTCACACGAGCGGTGAT 1143
 QY 409 -----LysAsnLeuGluSerLysLeuSerLysGlnGluSerValSerHisTh 424
 Db 1144 GGACAGCTGCAGCGCGCACCATGAGCAGAGCTGCAGCGGTGCAGCAGGAGAGGAGTG 1203
 QY 424 rLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyrAspLysAlaTy 444
 Db 1204 GCTCTGGCGGAGGAGCGGTGCCCGCTCAGCTTATTGAACCCATGAAGAGGCGTA 1263
 QY 444 r-----AsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArgAsnSe 462
 Db 1264 TCAAGAAGAGCTAAGCCGGAGCTTGACAAACCCCGGAGTCTCCAGCAGGCGCCAGAGAG 1323
 QY 462 r-----AspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAs 475
 Db 1324 CCTTCGAAGCAACACCGCTGGAGATGGAGCGCTAAAGCAGGAGCTGCAGGTGCTATC 1383
 QY 475 nAspGluSerThrAsnLys----- 481
 Db 1384 TGAGAGTACTCGCAGAAGTGCCTGGAGATCGGGCCCTGCACAGGCGAGGAGAGCG 1443
 QY 481 ----- 481
 Db 1444 TGACACACCTCGCGGCTGCCAGCAAGAGCGCCAGGAGTGTGCGCCCAACACAGGA 1503
 QY 482 -----GluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleTh 496
 Db 1504 GTTGCACTCCCACTGTCCGAGGAGATCGACCGCTCTCGCAGCTTCTACCTCGCGAGGG 1563
 QY 496 rHisProGluArgLeuGlyLysProAsn----- 505

Db 1564 CACAGGCAACAGCTGCGGGCGCAGCAATGAGAGGAGTTCCTGTGTGAGTGGAGGTCTCTCT 1623
 QY 506 -----SerGlnIleGluTyrThrGluAspG1 514
 Db 1624 CGAGTGAAGGAGAAATGAACCTCCAGTACCTGAAGAAGGAGGTGCAGTCTCTCCGAGATGA 1683
 QY 514 uValArgIleAlaGlnLeuAlaAspLysTyrThrSerAspGlyTyrIlePheAspG1 534
 Db 1684 GCTCCAGGTGATACAGAAGGACAAGCGCTTCACGGGAAGTACCAGGATGTGTACTGTGA 1743
 QY 534 uHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHi 554
 Db 1744 GCTGAACACATCAAGACTCGTGGAGCTGAGAGCTGAGCAGCTGAAGAGGACCTCTCG 1803
 QY 554 strIleGlyLysAspSerLeuSerAspLysGluLysVal----- 567
 Db 1804 C---CTGGCCATGGCTGCCCTGCAGGAGAGAGGCTGTGAGGAACAGCTTTGGCTGTGTA 1860
 QY 568 -----AlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuProPr 581
 Db 1861 GAGTCCCTCTGGTTCGATCAGCTGAGGAC-----CCTCC 1896
 QY 581 oSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyrAs 601
 Db 1897 AGCCCCA-----GGGACCCAGCCACCTCTCTCTCTGG 1929
 QY 601 nArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHisTh 621
 Db 1930 CAGAAGTCAGAAGCAAGCTTCTCCCC-----ACCCTGGGTCCATGTGCACCTCACAC 1983
 QY 621 rValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIleLy 641
 Db 1984 CACACACACC-----CACACACACACG----- 2005
 QY 641 sPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLe 661
 Db 2006 -----CACATACACACACACACACACACGATATGCACAC 2046
 QY 661 uPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAspG1 681
 Db 2047 ACATGCA-----CACACATACACGACACACACATGCACAGACA 2085
 QY 681 yTyrGlyAsn 684
 Db 2086 CTGGGGATCT 2095
 RESULT 9
 AK029960
 LOCUS
 DEFINITION
 Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4932411D20 product:similar to CTCL TUMOR ANTIGEN SE2-2 (FRAGMENT) [Homo sapiens], full insert sequence.
 AK029960
 VERSION
 AK029960.1 GI:26325891
 HTCC; CAP trapper.
 KEYWORDS
 Mus musculus (house mouse)
 SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 1
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
 20493374
 2
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 10349636
 99279253
 19-44 (1999)
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)

457 AsnLysGlyArgAsnSerPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAsp 476
 546 AAGATGGAATTCATCAAGAAATGATCAATGACAGACAGATTAAGAGACTGTCAGT 605
 477 Glu-----SerThrAsnLysGluLysLeuValAspAspLeuLeuAla 490
 606 CGACTGCAGAGCAAACTTTCATAGATAACAAGCAAGCTTTAATCGATGAACCT- 659
 491 PheLeuAlaProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyr 510
 660 -----CAAAGAAAGTTAAATAAATCTGAAAGCAACCACTGGAAGA 698
 511 ThrGluAspGluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyr 530
 699 AAGGTGATGATGATAGATTAAGCCGCTGAAGGAAAGAGTAGT- 743
 531 IlePheAspGluHisAspIleSerAspGluGlyAspAlaTyrValThr- 547
 744 -----AAAGAAGAATTAATTAGTGTGGAAGAGTAAGAATGGCAAAACCAAGTAGAG 797
 548 -----ProHisMetGlyHisSerHisTyrIleGlyLysAsp- 559
 798 GGACTACGAACACAGACTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 857
 560 -----SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyr 572
 858 AATACCTTAAAGGAACCTTTTGGCCAAAGCTGATAAGAGAACTTACTTTCAGAGAA 917
 573 ThrLysGluLysGlyIleLeuProSerProAspAlaAspValLysAlaAsnProThr 592
 918 CTGAAACCAACAGGAATG- 938
 593 GlyAspSerAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuVal 612
 939 GTTGACCAAGTTTGTAGAGT- -CGAGCTTGTGAATCTGAAAGAGAGTGTGAA- 989
 613 ArgLeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIlePro 632
 990 -----GAGCTAAAGAAAGAAATCTGAGACTGAGAA 1019
 633 HisLysAspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAla 652
 1020 AATGACATATTATACATGAGGACCCAGCAGCT- -CTT 1055
 653 ProAsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGlu- 670
 1056 CCAGAGATCTGTGTGGAAGACTTACATTTACAAATAAATACCTTCAAGAAAACTT 1115
 671 HisProAspGluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeu 690
 1116 CATACTTTAGAAAAAACTTTCAAGAGAGAAATATTCTCAGCTTTTGCCTTCAGAAATA 1175
 691 GlyLysLysAspHis- -SerGluAspProAsnLysAsnLysAlaAsp 706
 1176 GAGTCAGATGATCAGCTGTCAAAAGAAACAAGAACTTCAGAGGAAATTTCAAGTTGCA 1235
 707 GluGluProValGlu- -GluThrProAlaGluProGluValProGlnVal 722
 1236 TCTGAAACATCGAGCTGAAATTTCAACTTGAACAGCAAAATAAAGATTTGCCAAGACTA 1295
 723 GluThrGlu- -LysValGlu 728
 1296 AAGAATCAAGTGAAGATTTCAAGAAATGTGTGAATTTCTTAGAAGGAAACTGGAA 1355
 729 AlaGlnLeuLysGluAlaGluVal- - 736
 1356 CTGAGCGGAAGCTTGTGTCAGCTCAGAGGAGGCTGTAGAGGCTGGAGAGACAAATCCCGAA 1415
 737 -----LeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAla 750
 1416 CTGAAAAAACCATTTGGTTTAATGAAGAAATTAGTTGAAAGAGTCCAAAGAGAAAT- 1472
 751 ThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMet- 766

Db 1473 ---GAACAATTGAAAAGGCGATCAGGAATCTAGTACTAGTAAATAATGGCTACTATTGAG 1529
 QY 767 ---AspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780
 Db 1530 GAAGAAAATAGAAACTTAAAGGCTGAACCTAGAAAAGCTTAAAGCT 1574
 RESULT 10
 BC014691 3045 bp mRNA linear HTC 19-NOV-2003
 LOCUS Mus musculus protein tyrosine phosphatase, non-receptor type 21,
 DEFINITION mRNA [cdna clone IMAGE:4527498], containing frame-shift errors.
 ACCESSION BC014691
 VERSION BC014691.1 GI:15928418
 KEYWORDS HTC.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 3045)
 REFERENCE Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, W.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, M.J., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Mallek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettman, M., Maman, A., Rodriguez, S.,
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Butterfield, Y.S., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Dickerson, M.C., Krzywinski, M.I., Skalska, U., Smalish, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257
 12477932
 2 (bases 1 to 3045)
 REFERENCE Strausberg, R.
 AUTHORS Direct Submission
 TITLE Submitted (01-OCT-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
 COMMENT Contact: MGC help desk
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: The Cepko Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 cDNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: http://www.hgsc.bcm.tmc.edu/cdna/
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
 Kowles, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
 A.N., Gibbs, R.A.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
 Series: IRAC Plate: 31 Row: 0 Column: 6
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 6755239
 This clone has the following problem: frame shifted.
 FEATURES Location/Qualifiers
 1. .3045

/organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:4527498"
 /tissue_type="Eye, retina, mouse strain C57BL/6"
 /clone_lib="NTH MGC_94"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

ORIGIN

Alignment Scores:

Pred. No.: 3-54 Length: 3045
 Score: 139.50 Matches: 145
 Percent Similarity: 33.62% Conservative: 87
 Best Local Similarity: 21.01% Mismatches: 268
 Query Match: 3.35% Indels: 191
 DB: 11 Gaps: 33

US-09-765-271-56 (1-796) x BC014691 (1-3045)

QY 29 LysThrGluAsnLeuThrProAspGluValSerLysArgGluGlyLeuAsnAlaGluGln 48
 DB 1091 GAGGCGGAGCTCTGTACATGAGGAGGTAGAGAGATGGCTAGCGAGAGAGAGC 1150
 QY 49 IleValIleLysIleThrAspGlnGlyTyrValThrSer-----HisGly 63
 DB 1151 TACCCTGGCAAG---GACAGCCAGAGAGTGCATATCCATTGGTGCCTGTCTCATGGC 1207
 QY 64 AspHisTyrHisTyrAsnGlyLysValProTyr----- 75
 DB 1208 ATCTTTGTGAACACAGAAAGGAGGCTCTGTGGTGTTCAGGTGGCAGCATTTGCT 1267
 QY 76 -----AspIleIleSerGluGluLeuLeuMetLysAspProAsnTyr 90
 DB 1268 AACATGTGCAVACAAAGTCCTTTTTCAGTACAGAGCTGGCAATAAGAGAGACCATC 1327
 QY 91 LysLeuLysAspGluAspIleValAsnGluValLysGlyGlyTyrValIleLysValAsp 110
 DB 1328 CAGTTTCAAACTGAAGACATGGAGACA----- 1354
 QY 111 GlyLysTyrTyrValTyrLeuLysAspAlaAlaHisAlaAspAsnValArgThrLysGlu 130
 DB 1355 GCCAAGTACGTGGGGCTGTGTGGCAGCGCAAAATTTACAGACTGAACACAGTGT 1414
 QY 131 GluIleAsnArgGlnLysGlnGluHisSerGlnHisArgGluGlyGlyThrProArgAsn 150
 DB 1415 AGCCTGCAACACGAGCGCCACGTCGTAACACTCAGTCAGAGGGGCTCATCTCCAGG--- 1471
 QY 151 AspGlyAlaValAlaLeuAlaArgSerGln----- 160
 DB 1472 -----ATGTCCTGCTAAACCCAGCCCTATGCGATGCCCTCCACCCAGCTG 1522
 QY 161 -----GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIlelle 177
 DB 1523 CATTTATAATGGACATTATACA-----GAGCCATT 1552
 QY 178 GluAspThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLys 197
 DB 1553 GCTTCTTCCCAAGACAAACATCTTTGTGCCCAAGAAATGGATTCTACTGTCACTCCAG 1612
 QY 198 AsnGluLeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsn 217
 DB 1613 ACNAGCTGATAGAACCCAGATTGAC-----CTCAGTGGTGC----- 1651
 QY 218 LeuSerAsnSerArgThrTyr-----ArgArgGlnAsnSerAspAsnThrSerArgThr 235
 DB 1652 ATCCGCAACGCGCAGCTCTACAGTGCACAGTACGAATTCCTTAAACACACTTCAGCCC 1711
 QY 236 AsnTyrValProSer-----ValSerAsnProGlyThrThrAsnThrAsnThrSerAsn 253
 DB 1712 TACCTGAGCCCTCCCTCCATGTCCTCCACCCGAGTATCCTGGCAGTGTATCATGAGA 1771
 QY 254 AsnSerAsnThrAsnSerGlnAlaSerGln-----Ser 264

DB 1772 CCCGACTACATCCCGTCCACCGGCACAGTGGCTCATTTCCCGCGTCTTACCGCCGACC 1831
 QY 265 AsnIlePheAspSerLeuLeuLysGlnLeuTyrLys---LeuProLeuSerGlnArgHis 283
 DB 1832 CTTGACTACGAGACCGTGTATGAGCAGCTCAACAGAGCATGGTGACGACGACCGGAC 1891
 QY 284 ValGluSer-----AspGly 288
 DB 1892 AGCCACTCGTGAGGAACCTCAATTGGCAGTTCTCTATGCGTACAGCGGCCGACGCC 1951
 QY 289 LeuValPheAspProAlaGlnIleThrSerArgThrAlaArgGlyValAlaValProHis 308
 DB 1952 TTGTCTTACAGCCAGCCGAGATT-----CGGAGACCCCTCACTCACTCTCCCGAG 2005
 QY 309 GlyAspHisTyrHisPhe---IleProTyrSerGlnMetSerGlu-LeuGluGluArgI 327
 DB 2006 TCAGCCCACTACCCCTTTAACTGAACCTACAGTTCCTCCAGCCAGTCTC----- 2054
 QY 327 eAlaArgIleIleProLeuArgTyrArgSerAsnHisTyrValProAspSerArgProI 347
 DB 2055 ---CGTATCCCTACCTGCCGAGAGCGCCAGTGGTGGGCGCGTG-AGTGTGCTGA 2109
 QY 347 u-----GlnProSerProGlnProThr----- 354
 DB 2110 ACTGACAAATGTGCAGTGCAGGCTCAGGACTACCCAGCTCCAAACATTTATGAGAACCCA 2169
 QY 355 -----ProGluProSerProGlyProGlnProAla-----ProAsnIle 367
 DB 2170 GGTGTACCGCCACCCCGCGGTATCTTTACCAAGCTTGCCTCAACAGCACCCGAGACT 2229
 QY 367 uLys-----IleAspSerAsnSerSerLeuValSerGlnLeuValArgLys 382
 DB 2230 GTCCCGCCACCTTACATCAGCAGCAGCAATCCGATCTCATCACAGCGCGGTCCATCA 2289
 QY 382 sValGlyGluGlyTyrValPheGluLysGlyIleSerArgTyrValPheAlaLysAs 402
 DB 2290 CTCAGTGCAG-----ACCTTCAGAGGAGACACTTGCCTGGCTCACTCTCTGAGGA 2343
 QY 402 pleuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSerVal-- 421
 DB 2344 GGTACGAGCGCCCTCAGCAGCGCGCCATCCCACTTACAGAGAGAACATGATTGA 2403
 QY 422 -----SerHisThrLeuThrAl 427
 DB 2404 ATTCGAGGGCTCACACATGGATTTCGAGGCTCAGGCTCAAGGAGAGACGGTGTGTCAGC 2463
 QY 427 aLysLysGluAsnValAlaProArg-----AspGlnGluPheTyr 440
 DB 2464 CTCAGCAGCAGACGTGGCGCCCGGACCTTCTCAGCAGGCTCCCGAGTCCAGCGTCTTCTC 2523
 QY 440 rAspLysAlaTyrAsnLeuLeuThrGluAlaHisLys---AlaLeuPhe**AsnLysG 459
 DB 2524 TGACAAATGAAGCAGGAGGGGACTGACAGCAGCAGAGGGTGCAGATACAGCCATAGA 2583
 QY 459 yArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAsp-----G 477
 DB 2584 GTCCCTTTCTGATGCCACCATCTGATTTCATGAGTGTGAGGAGGACGAGACTTGAAGA 2643
 QY 477 uSerThrAsnLysGluLysLeuValAspLeu----- 488
 DB 2644 GGACAGCAGCAGAGAAGACGGCCATCTCAGCTGTGTGAGCCCCCGCTTACGGCGCGCTT 2703
 QY 489 -----LeuAlaPheLeuAlaProIleThrHisProGluArgLe 501
 DB 2704 CTCTCAGGAACCTAAACTACCTCCCTGTGCTTCAGGACTCCGATCTCGAGCGCTCTGCATAT 2763
 QY 501 uGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeuAl 521
 DB 2764 TTTTGAGCCCAAGCCCATGTC-----ACAGAGCTGTGAGAGAGGCGGCAAGACATCAG 2817
 QY 521 aAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAspG 541

Db 1352 -----CGAAGAGCCCTGCAAGACAATCACACCACAGAAAT-----GCT 1390
Qy 434 ProArgAspGluGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453
Db 1391 CCAAGAGAGGAC-----TCTAGGGGGCGTCCCTCT 1420
Qy 454 LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473
Db 1421 TTTTATCTCTCGGAGGAGAGATCAGAAACAGCAAGAGAGAAAGTTTGTAAAGG 1480
Qy 474 LeuAsnAspGluSerThrAsnLysGluLysLeuVal-AspAspLeuLeuAlaPheLeuAl 493
Db 1481 TTTACAGATGAA-----CAGTCTAGAGTATTCTCTGTAGAGGGGTAAATACC 1528
Qy 493 aProLeuThrHisProGluArgLeuGlyLysProAsn-----SerGlnIleGluTyrTh 511
Db 1529 AGGGATAAGAGGCCCAAGAGAGAAAGGGTCCGAGAAAGGCAGACAGCGAGATTGG 1588
Qy 511 rGluAspGluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrI 531
Db 1589 -GACATCAGAGGTTCTAGATTACTTTAGTATAAGAGCTCGCAAAACAAAGCTTC-- 1645
Qy 531 ePheAspGluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetG 551
Db 1646 -----CATGAT-----TCTCAAGGGGATGAC-----ACAGAGGAGACAGA 1680
Qy 551 yHisSerHisTrpIleGlyLysAspSerLeuSerAspLysGluLys-----ValAl 568
Db 1681 GGATTATAGACATTTAGGAATCTAGTCTGCGAGATCAGGGGAAGAGCTTTGCTACTTC 1740
Qy 568 aAlaGlnAlaTyrThrLysGluLysGlyIleLeuProProSerProAspAlaAspVally 588
Db 1741 ATCTCACCAGGATATCTAGGAGGAAGGA-----CCCAAGTACAGTCCAAAGTTTC 1791
Qy 588 sAlaAsnProThrGlyAspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysAr 608
Db 1792 ACTAAAAGGCAGTAGAGAAAGT-----GATGGATTATAGAGAAAGAAAAA 1836
Qy 608 gileProLeuValArgLeuProTyrMetValGluHisThrValGluValLysAsnGlyAs 628
Db 1837 TTATAAACTGAAAGAGACTGTCATAGTAGTGAAGGCCCTAGACTACAAAGAC----- 1891
Qy 628 nLeuIleLeuProHisLysAspHisTyrHisAsnIleLysPheAlaTyrPheAspPhi 648
Db 1892 -----AGCACAGGAC----- 1903
Qy 648 sThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAla-----ThrI 665
Db 1904 -----GAAAGACAAAGGCTCTGACAGAAATACAGT 1932
Qy 665 elysTyrTyrValGluHisProAspGluArgProHisSerAsnAspGlyTyrGlyAsnAl 685
Db 1933 GAAGAAAGAGTGCAGTCCCGGAGCAG-----GTAAA 1965
Qy 685 aSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsnLysAsnPheLysAl 705
Db 1966 GTCTGAAAGACTCAAAGAGCTCTTTGATTACAGTCCCTCTACACAGAGCTGACAGCG 2025
Qy 705 aAspGluGluProVal-----GluGluThrProAlaGluProGluValProGlnValG 723
Db 2026 ACCAGAAAATCCATCTTCAGAGAGAGAGCCCGCTGAGGATCAAAATGATACCCAGTGA 2085
Qy 723 uThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuAlaLysValThrAs 743
Db 2086 TTCTCAGCTCTGAGAGTCAAACTCAAAATGGCCCTGTCCTCTG-----GATGA 2136
Qy 743 pSerSerLeuLysAlaAsnAlaThrGlu 752
Db 2137 CTCACACAGACCTGCCTCTCTGACTAAA 2164
RESULT 12
AY404467 3573 bp DNA linear GSS 12-DEC-2003
LOCUS

DEFINITION Pan troglodytes HCM1900 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY404467
VERSION AY404467.1 GI:39760444
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
REFERENCE 1 (bases 1 to 3573)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 3573)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..3573
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>3573
/locus_tag="HCM1900"
gene
ORIGIN
Alignment Scores:
Pred. No.: 4.95 Length: 3573
Score: 139.00 Matches: 182
Percent Similarity: 34.57% Conservative: 136
Best Local Similarity: 19.78% Mismatches: 350
Query Match: 3.34% Indels: 254
DB: 29 Gaps: 40
US-09-765-271-56 (1-796) x AY404467 (1-3573)
Qy 31 GluAsnLeuThrProAspGluValSerLysArgGluGlyLeuAsnAlaGluGlnIleVal 50
Db 7 GAGGAGGATACCAAGAGAGAGTCCCAAGAGAGATGGAGTT----- 48
Qy 51 IleLysIleThrAspGlnGlyTyrValThrSerHisGlyAspHisTyrHisTyrTyrAsn 70
Db 49 -----GGTGATGCAACATTTCGNNNC 72
Qy 71 GlyLysValProTyrAsp---AlaIleIleSerGluGluLeuLeuMetLysAspProAsn 89
Db 73 AGCAGATATGGAAGAGCATAGCTACATACAAAGAGGATGATTACAA-----AA 126
Qy 90 TyrLysLeuLysAspGluAspIleValAsnGluValLysGlyGlyTyrValIleLysVal 109
Db 127 GTAGCTTAGAAGAGACTCTGTGGTGCAATCTGTTGAGAACGATTCCTCCAAACATGATG 186
Qy 110 AspGly-----LysTyrTyrValTyrLeuLysAspAlaAlaHisAspAsn 125
Db 187 GAGAGCCTCAGCCCAAGAAATACTCTCCAGTCTGAGATTTAAAGCAATGAGACTAT 246
Qy 126 ValArgThrLysGluGluIleAsnArgGln---LysGlnGluHisSerGlnHisArgGlu 144
Db 247 TCTGGCTCTCTATTATACCTCTCACACCTGTGCTGCAAGAGAGAGCCCTCTCTCTT 306
Qy 145 GlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaAspSerGlnGly---ArgTyr 163

Db 307 GGAACCACTGTTCAGAACG-----AGCCCTCTCTTAGCCAAATCCAGGAGCAAGCAG 360
 QY 164 ThrThrAspAspGlyTyrIlePheAsn-----AlaSerAspIleIleGlu 178
 Db 361 TTCTCTTATGATGGAACCTGACAAAATATTCCTATGAAACCTTCAAACTCTTACTCAGC 420
 QY 179 AspThrGly-----AspAlaTyrIleValProHisGlyAspHisTyrHisTyrIlePro 196
 Db 421 ACTACATCTCTCCCTCAGTGGATATCCACTTGGAGAGCAGACTTTGATCATTTACTGCGC 480
 QY 197 LysAsnGluLeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGly 216
 Db 481 CGGAC-----AGTCAAGGCGCTTG----- 501
 QY 217 AsnLeuSerAsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsn 236
 Db 502 AGGCTCTCAGAGAGCCTCCCTATTCCAAATATAGCTCAAGGAATAAATCGACAAAT 561
 QY 237 ----- 561
 Db 562 GTCTACTTTCTTGGAGGCTCGAAGGTGGAAGGCATCTGCTCGCTCGCTCGCCATGTGG 621
 QY 238 -----ValProSerValSerAsnProGlyThrThrAsnThrAsnThrSerAsn 253
 Db 622 AATGGAAGTTCCCTCAGTGTATGCTGGCCGCTCTCTATCAGCAGATCGGAGCGCAAGC 681
 QY 254 AsnSerAsnThrAsnSerGlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGln 273
 Db 682 ATGCTTTCAAGCCCAAGCAGCAGGAAATGAGCATTCAGGACGAGCTCAGCGCTTCAA 741
 QY 274 LeuTyrLysLeuProLeuSerGlnArgHisValGluSerAspGlyLeuValPheAspPro 293
 Db 742 -----CCCAAGTTAACTAGACACAGGAG-----CTTGCATCTGAAAC 780
 QY 294 AlaGlnIleThrSerArgThrAlaArgGlyValAlaValProHisGlyAspHisTyrHis 313
 Db 781 ATCAATTTGACAACTAGGAAGTACTCCAGCAGCAGCCTGAGTCACATGGAGCCTAC--- 837
 QY 314 PheIleProTyrSerGlnMetSerGluLeuGluGluArgIleAlaArgIleIleProLeu 333
 Db 838 -----AGCCGATCACTTCCCGAGG 855
 QY 334 ArgTyrArgSer-----AsnHisTrpValProAspSerArgProGluGlnProSerPro 351
 Db 856 TTGTACAGAGCCACAGAGAACCTGACACTCTCAGCTTGCTTCCAGAAACTCTCTG 915
 QY 352 GlnProThrProGluProSerProGlyProGlnProAlaProAsnLeuLysIleAsp--- 370
 Db 916 GGCAATTTCCAAACGAAACAACTTTGGGAAAGGATCTACCTCATAGCGTAATAGACAAT 975
 QY 371 -----SerAsnSerSerLeuValSerGlnLeuValArg 381
 Db 976 GACAAATTACCTTAATTTTCTTTGAGCTCAGGGGCTTTACCTATAAAACCTCTGCT 1035
 QY 382 LysValGluGluGlyTyrValPheGluGluLysGlyIle-----SerArgTyrValPhe 399
 Db 1036 TCTGAAGGCAATCTCTTATGTAAGTTCTACCTCAGTCTGCTGCGCAGTCCACGAGTGGCT 1095
 QY 400 AlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGlu 419
 Db 1096 CGGAAGATGCTTCTGGGCTCCACCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1155
 QY 420 SerVal-----SerHisThrLeuThr-----Ala 427
 Db 1156 TATGTGGGGCAAAACCCGAGTCAATTAATTTCTGCTGGGAGTCCAGACAGATTTTGGG 1215
 QY 428 LysLysGluAsnValAlaProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeu 447
 Db 1216 ACCAGGAGGAACTCTCTTGTGGATCTGTGGAATTTGATGAGGCTGATTTGGAAAGCCTC 1275
 QY 448 ThrGluAla-----HisLysAlaLeuPhe***AsnLysGlyArgAsnSerAsp 463
 Db 1276 AGCAGGCTCAGGAACCCCGCCCTCTCGGAAACCGGAAAGCAGTATTAGTCTCC 1335

QY 464 PheGlnAlaLeuAspLysLeuLeu-----GluArgLeuAsn 475
 Db 1336 ATTTCCAGCAGCTGTGACCTGATGATTATCACCAGCGGAGAGGAGGAGGAGGAGGAGGAGG 1395
 QY 476 AspGluSerThrAsnLys-----GluLysLeuValAspAspLeuLeuAlaPheLeuAla 493
 Db 1396 GAGCAGGAAATGAGCGGATTTGGAGACAGCGCTTGGAGACCATCTCAGTCTCTGTGCT 1455
 QY 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGlu-----Tyr 510
 Db 1456 GAATACACA-----AAGCTGACAGCTCGCTTATCTACTGGGACCACC 1497
 QY 511 ThrGluAspGluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyr 530
 Db 1498 GTGGAAGATGTGCAGAAAATCAACAAGAGAGCTTGAAGAGTGCAGCTCTCTGTAGGAG 1557
 QY 531 IlePheAspGluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMet 550
 Db 1558 TCTGTGTTTGAAGAGCCCTCATGAGCCCTGAC-----ACAAGATACAGG 1602
 QY 551 GlyHisSerHisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAla--G 570
 Db 1603 TGCCACCGG-----AAAGACTCCCTCCTGATGACAGCTTGCAGAGCTGTGGG 1650
 QY 570 IlnAlaTyrThrLysGluLysGlyIleLeuProProSerProAspAlaAsp-----ValL 588
 Db 1651 AGTCTCAGTTCAGACAGTGCAGCTTCTTACCCCGCAGGAGCAGCAGGAGATGATGAAC 1710
 QY 588 ysAlaAsnProThrGlyAspSerAlaAlaIleTyrAsnArgValLysGlyGluLysA 608
 Db 1711 CTGAGCCAGCTCACCCGAGCTCTCCACCACTCTCCACCTTCCGAAAGCTTCCAGC 1770
 QY 608 rgIleProLeuValArgLeuProTyrMetValGluHisThrValGluValLysAsnGlyA 628
 Db 1771 GAGT-CCTTATTAAGTATCTTACCAAGACCCAGAGGTTATAGTGAAGAACAGAG 1829
 QY 628 snLeuIleIleProHisLysAspHisTyrHisAsnIleLysPheAlaTrpPheAspAsp 648
 Db 1830 ATCTC-----AGGAGTTGGCTGCAA--- 1849
 QY 648 isThrTyrLysAlaProAsnGlyTyr-----ThrLeuGluAspLeuPheAlaThrI 665
 Db 1850 -----TGAAGAAACCCGGATA-GTCAATCTGAACAACTCGAGGAACTTAAGCAAAA 1903
 QY 665 lelystTyrTyrValGluHisProAspGluArg----- 675
 Db 1904 TCAAGACATAAATGATCAGATGATGAGTCTTTCAGAGAGTGGATATGGAATGTGCTC 1963
 QY 676 -----P 676
 Db 1964 TTTTGGATGAGAACAGAAAATCTGAAACAACTTGAACCTTATGAAGGAGAGAGATTTGG 2023
 QY 676 roHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHis 696
 Db 2024 ATCATCTAAACCCGAAATAGCTGAACCTGGAAGAACATTTGTTGTAAGAACCAAGG 2083
 QY 696 erGluAsp-----ProAsnLysAsnPheLysAlaAspGluGluProValGluG 712
 Db 2084 ATGCTGACCTGTTGGATGTTGAAAGCAACACATTTGAA----- 2121
 QY 712 luThrProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGlnLeu 732
 Db 2122 -----GACCTGGAGTTCCAGCAGCTTGAACATGAGAGCCGCTTAGATGAAGAAA 2170
 QY 732 ysGlu-----AlaGluValLeuLeuAlaLysValThrAsp----- 743
 Db 2171 AGGAGAACTTGACTCAACAGCTCTCTCGTGAAGTGTCTGAATATCAACGGAACATCGTTT 2230
 QY 744 -----SerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeu- 757
 Db 2231 CTAGAAAGGAAATAATTTCTGCAATTCGAAAGCAAGCAACCAATCATATTGTTTCAGCAGCTC 2290

```

QY 758 -----ArgAsnLeuThrLeuGlnLeuMetAspA 768
DB 2291 AGAGAGAACAGATCATTTTGTGAAGAAAGAAATTAATAATGATGTTGCAAGAG 2350
QY 768 snAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer----- 785
DB 2351 AAAGAGAGAACTTTGTAATTTGGAAGAAATACTCCAGCCTCTCTGGGGGAAAGGT 2410
QY 786 -----AsnProSerSer-----ValSerLysGluLysIleAsn 796
DB 2411 TTCCCGTTAAACCCCAATACTTTTAAAGAGGGGTATATCATGCTGTAATAATGAGATTAAT 2466

RESULT 13
CG756440/c 1831 bp DNA linear GSS 24-OCT-2003
LOCUS P051-3-G10.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
DEFINITION genomic survey sequence.
ACCESSION CG756440
VERSION CG756440.1 GI:37984016
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
AUTHORS Neodiplogasteridae; Pristionchus.
TITLES Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
Buntjer,J., van der Meulen,M. and Sommer,R.J.
An integrated physical and genetic map of the nematode Pristionchus
pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE 22835951
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.

FEATURES
Location/Qualifiers
1..1831
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of
the genomic DNA with EcoRI and cloning into the BAC
vector."

ORIGIN
Alignment Scores:
Pred. No.: 1.98 Length: 1831
Score: 138.50 Matches: 135
Percent Similarity: 31.09% Conservative: 87
Best local Similarity: 18.91% Mismatches: 239
Query Match: 3.33% Indels: 257
DB: 29 Gaps: 28

US-09-765-271-56 (1-796) x CG756440 (1-1831)
QY 133 AsnArgGlnLysGlnGluHisSerGlnHisArgGluGlyThrPro----- 148
DB 1765 AACGCACACACCGCAAGACACGCCCTCCCGACCGACGCGCGCCGCCGACACCG 1706
QY 149 -----ArgAsnAspGlyAlaValAlaLeuAla 157
DB 1705 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1646
QY 158 ArgSerGlnGlyArgTyThrThrAspAspGlyTyThrIlePheAsnAlaSerAspIle 177
DB 1645 CGCGCGCAACCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1586

```

```

QY 178 GluAsp-----ThrGlyAspAlaTyr-----IleValPro-----HisGly 189
DB 1585 TRATCAATAGCAACAGGACAGACACAAACGATACATACCGCCGCGAGACACACGCT 1526
QY 190 AspHis-----TyrHisTyIleProLysAsnGlnLeuSerAla 202
DB 1525 AAACACGCGCACCAAAAAACGACGACACACACAGAAACACATAACAA-ATACAGACC 1467
QY 203 SerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSerArg 222
DB 1466 GAAACACCTTAGGACAGACGAC-----CGACCACTCGCAACAAACACCA 1422
QY 223 ThrTyArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSerValSer 242
DB 1421 AACACGCGACACGCAACAGCGAGACTAGCAGAAC-ACAAATNCCCACGCGGTACGA 1363
QY 243 AsnProGlyThrThrAsnThrSerAsn-AsnSerAsnThrAsnSerGlnAlaSe 262
DB 1362 ACAAGGCCAACGCCCAACACGAAACACAAACCGCAGCGCACGCAACAAAGTAGA 1303
QY 262 rGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyLysLeuProLeuSerGlnAr 282
DB 1302 AAAGCACAAAGAG-----AGAAAGAGCAGACACAAAGACAGCAAAACACCA 1255
QY 282 GHIS-----ValGluSerAspGlyLe 289
DB 1254 GCACAAACACACGACACACATCAACCATAGAAACACACCGCAGACGACCAACAC 1195
QY 289 uValPheAspProAlaGlnIleThrSerArgThrAlaArgGlyValAlaValProHisGl 309
DB 1194 AGCGCCACCCACACACAGGAGCAGTGCGCC----- 1161
QY 309 yAspHisTyHisPheIleProTySerGlnMetSerGluLeuGluGluArgIleAlaAr 329
DB 1160 -----CCAACGCACACACCGCGCAACCGCAACCAACCAACCAACAG 1120
QY 329 gIleIleProLeuArgTyArgSerAsnHisTrpValProAspSerArgProGluGlnPr 349
DB 1119 AGAGCACACCCACACACACAG-----CACACACGAACAAACACGACGACACACCC 1066
QY 349 oSerProGlnProThrProGluProSerPro-----GlyProGlnProAlaProAs 366
DB 1065 ACACCCACGACAAACAAACACACACCCCGCCCATACAAACCGGACCCCAATGCCACCA 1008
QY 366 nLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuValArgLysValGlyGluGl 386
DB 1007 -----GCATATCCACACACAGCATGGCACCACAC----- 978
QY 386 yTyValPheGluGluLysGlyIleSerArgTyValPheAlaLysAspLeuProSerGl 406
DB 977 -----ACACAAGAGACACACGACACCC 958
QY 406 uThrValLysAsnLeuGluSerLys-LeuSerLysGlnGluSerVal----- 421
DB 957 CACCAACAAAGAACACGAAACACAGCCGCGACCGACCGAAGAAACACACACACACGCG 898
QY 422 -----SerHisThrLeuThrAlaL 428
DB 897 GACCCAAACACCCACACCGCAACAAACACTCNCCCACACACACACACACACACAC 838
QY 428 yLysGluAsnValAlaProArgAspGlnGluPheTyArgLysAlaTyAsnLeuLeuT 448
DB 837 GCACACCCCAATCGACCGCCCGCCACACACCGCTCGCAGACGACGACCCACACATTCAC 778
QY 448 hrGluAlaHisLysAlaLeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuA 468
DB 777 CGCACGCGCACACGAGCA----- 761
QY 468 sPylsLeuLeuGluArgLeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspL 488
DB 760 -----CAACACACACACAGAGGAAACAA----- 737

```



```

/db_xref="MGI:2392424"
/db_xref="taxon:10090"
/clone="4931417E02"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
229..3147
/notes="unnamed protein product; putative
retininitis pigmentosa GTPase regulator (MGI:1344037,
GB|NM_011285, evidence: BLASTN, 99%, match=1957)"
/codon_start=1
/protein_id="BAC26649.1"
/db_xref="GI:26325790"
/translation="MPGRSWSQGVGHLRLNRVAPALFPKQAQIPFAGMAESES
LVPDGVAFVFTKTPAFNIPKFWKNDIPICLSCGDEHTAIVTGNKGLYFGNNW
GQLGLGSAKAIKPTCIKALKEPKVLAACGRNHTLVSTDTGVYAAGNNEGOLGLG
DTRDRTFHQIVFPTADTKQLSAGANTSAALTEGDKLFWMGDNSEGIQIULEKSNV
CIPHEVTQKPTSMISCGYHSFAFVMDGELYTFGEPEGKLGLEPGLNLMHRSQVR
LGIPEVIOVACGGGHTVLTBKVVAVAGLQFGOLGLTFLFETSEPKIIRIKDQK
ICHISGHNHTALTEGLLTYTFGGRHGKGLGLENFTNOFPPTLCSNLFRLFVOLI
ACGGCHMLVAFVPTIDEPKFDVIEPIYISTGFSINDLSRSLRSLSERLRRR
ERERPPCSASMTGLPPELGTSASTSAVPSPFPHLSVNNYPSKSPSESLEPDDSD
YFEDKXNDKTETENSADVSENFETDILNTHMTTSSNEKLLDFSPQIQOQSDT
PEKVMESTPCTENSDSYEEEMSKIKEVTYKQYLAAGLYMIRPAEILAEAFDSRQVN
GLDQVEERVFDTGKLGLOSKQVSKESDEIVSEKKEVMEVADVKKIRESENSKSDS
LFPDDPKVMSESDNKDIABERSSEQNMTFDTSETELVEPDSYMECERHSEQDSA
LELEQPKLVYESSEKDEKDDDEVENTEWYDRNCTEQETENVFRAIRFFPKFDL
KHDSLGIPPEEGDEEGNVVQVQAQENLEFEGDKERAKAPSDVITKEBA
POLSETVPEKEBEMBEISILNVEDTVEERKEGEKEIVEGSIPETEGSETIDITDE
KLDEVLKEDSDASLLQALREYNENPKMHYDRVKSSSEILGNDPTSKDIKKAKKS
3291..3296
/notes="putative"
3313
/notes="putative"
ORIGIN
Alignment Scores:
Pred. No.: 5.33 Length: 3313
Score: 138.00 Matches: 131
Percent Similarity: 33.29% Conservative: 107
Best Local Similarity: 18.32% Mismatches: 248
Query Match: 3.31% Indels: 229
DB: 11 Gaps: 28
US-09-765-271-56 (1-796) x AK029864 (1-3313)
QY 108 LysValAspGlyLysTyrValTyrLeuLysAspAlaAlaHisAlaAspAsn---Val 126
Db 1468 MAATTGAGACGTATATGACGCTTATATAGTACAGGTCTTTTCCATCAATGACCTC 1527
QY 127 ArgThrGluGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 146
Db 1528 TCCCCAAGAAGTTCACTGAATAGATCTTTATCAGCACGCTCGCGCGAAGAGCGGGAG 1587
QY 147 ThrProArgAsnAspGlyAlaValAlaAlaLeuAlaArgSerGlnGlyArgTyrThrAsp 166
Db 1588 AGACCCCCAGCTGCTCAGCTTCAATGGTG-----GGAACACTGCCTCCATTA 1632
QY 167 AspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThrGlyAspAlaTyrIleVal 186
Db 1633 GAGGGGACTTCTGCCCTCCACTTCA-----GCTTATTTTTC 1668
QY 187 ProHisGlyAspHisTyrHis-----TyrIleProLysAsnGluLeuSer 201
Db 1669 CCCAGTTTCAACCCCTTCCATTTGCTGTGAATAACTTACCACAGAGAAAGCCCTCTGAA 1728
QY 202 AlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSer 221
Db 1729 TCAATGAGCCACTGGACTAGATATTTTGAAGATAAATGAACAAGACACAGACACA 1788
QY 222 ArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSerVal 241

```

CDS

polyA_signal

polyA_site

ORIGIN

```

1789 GAAATTCTTCAGCAGTGGATTACAAAAACCTTTGGTGAACCTAATATCTTAATATG 1848
Db 1789 GAAATTCTTCAGCAGTGGATTACAAAAACCTTTGGTGAACCTAATATCTTAATATG 1848
QY 242 SerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAla 261
Db 242 SerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAla 261
QY 1849 ACACATATGATCAGTACACGATCCCAATGAGAGTTATTAGATTTTACCAATTCACAAA 1908
Db 1849 ACACATATGATCAGTACACGATCCCAATGAGAGTTATTAGATTTTACCAATTCACAAA 1908
QY 262 SerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSerGln 281
Db 262 SerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSerGln 281
QY 1909 CACACAGATCAA-----GACACATTTGAGAAAGTGAGAAAGTACACCGTGCACTGAA 1962
Db 1909 CACACAGATCAA-----GACACATTTGAGAAAGTGAGAAAGTACACCGTGCACTGAA 1962
QY 282 ArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThrAla 301
Db 282 ArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThrAla 301
QY 1963 AATGAG----- 1968
Db 1963 AATGAG----- 1968
QY 302 ArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMetSer 321
Db 302 ArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMetSer 321
QY 1969 -----GATAGTTATGAA-----TATGACAAAGATGCA 1995
Db 1969 -----GATAGTTATGAA-----TATGACAAAGATGCA 1995
QY 322 GluLeuGluGlu-----ArgIleAlaArgIleIleProLeuArgTyr 335
Db 322 GluLeuGluGlu-----ArgIleAlaArgIleIleProLeuArgTyr 335
QY 1996 AAAAAAAGAGTGCACAGTGACAAACAATATTTAGCAAAAGGATCTACATG----- 2049
Db 1996 AAAAAAAGAGTGCACAGTGACAAACAATATTTAGCAAAAGGATCTACATG----- 2049
QY 336 ArgSerAsnHisTrpValProAspSerArgProGluGlnProSerProGlnProThrPro 355
Db 336 ArgSerAsnHisTrpValProAspSerArgProGluGlnProSerProGlnProThrPro 355
QY 2049 ----- 2049
Db 2049 ----- 2049
QY 356 GluProSerProGlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeu 375
Db 356 GluProSerProGlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeu 375
QY 2050 -----ATACGACACCCGAGATCTGGAAGCATTTTCAGATGAGNA--- 2091
Db 2050 -----ATACGACACCCGAGATCTGGAAGCATTTTCAGATGAGNA--- 2091
QY 376 ValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSer 395
Db 376 ValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSer 395
QY 2092 GTGGGTAAATGGCTAGACCCAGTGGAGGCGCGGTCTTCTCCTGATGGA----- 2142
Db 2092 GTGGGTAAATGGCTAGACCCAGTGGAGGCGCGGTCTTCTCCTGATGGA----- 2142
QY 396 ArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeu 415
Db 396 ArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeu 415
QY 2143 -----AAGGGCTTACAAAGCAAGCAA 2163
Db 2143 -----AAGGGCTTACAAAGCAAGCAA 2163
QY 416 SerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArg 435
Db 416 SerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArg 435
QY 2164 GTAGGAAGGAAGTATGATGAAGAAATAGTCTCTGAGAAAGAACTCAGGTG----- 2214
Db 2164 GTAGGAAGGAAGTATGATGAAGAAATAGTCTCTGAGAAAGAACTCAGGTG----- 2214
QY 436 AspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe 455
Db 436 AspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe 455
QY 2214 ----- 2214
Db 2214 ----- 2214
QY 456 ***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsn 475
Db 456 ***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsn 475
QY 2215 -----ATGGAAGTGGCAGATGTGAAAAGATTTAGAGAAAGT 2250
Db 2215 -----ATGGAAGTGGCAGATGTGAAAAGATTTAGAGAAAGT 2250
QY 476 AspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIle 495
Db 476 AspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIle 495
QY 2251 GAGAGAACTCAAAATCAGATTCACCTTTTGTGACTTA----- 2289
Db 2251 GAGAGAACTCAAAATCAGATTCACCTTTTGTGACTTA----- 2289
QY 496 ThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluVal 515
Db 496 ThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluVal 515
QY 2290 -----CCAGATAAA-----ACCATCAATTTCTGAG-----AGTGAAGCAATAAA 2328
Db 2290 -----CCAGATAAA-----ACCATCAATTTCTGAG-----AGTGAAGCAATAAA 2328
QY 516 ArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHis 535
Db 516 ArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHis 535
QY 2329 GATATTGCTGAG-----GAAAGGAGAGCAGTGCAGAGATATGACCTTTGACAGTGAA 2382
Db 2329 GATATTGCTGAG-----GAAAGGAGAGCAGTGCAGAGATATGACCTTTGACAGTGAA 2382
QY 536 AspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrp 555
Db 536 AspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrp 555
QY 2383 ACAGAAATGGTAGAAGAACCCAGACAGTACATGGAATGTGAGAGG---CACACTGAG--- 2436
Db 2383 ACAGAAATGGTAGAAGAACCCAGACAGTACATGGAATGTGAGAGG---CACACTGAG--- 2436
QY 556 IleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGlu 575
Db 556 IleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGlu 575
QY 2437 -----CAAGACAGTCTGAGGAGTGGAGCACCCCAAGCTAGTAGAATATAGTAGGAA 2490
Db 2437 -----CAAGACAGTCTGAGGAGTGGAGCACCCCAAGCTAGTAGAATATAGTAGGAA 2490
QY 576 LysGlyIleLeuProSerProAspAlaAspValLysAlaAsnProThrGlyAspSer 595
Db 576 LysGlyIleLeuProSerProAspAlaAspValLysAlaAsnProThrGlyAspSer 595

```



```

/dev stage="adult"
76_->2581
/Note="nonhistone chromosomal protein HMG-1 homolog [Bos
primigenius taurus] (PIR|S01947, evidence: FASTV, 89%ID,
97.6%length, match=625)
putative"

misc_feature
ORIGIN
Alignment Scores:
Pred. No.: 4.41 Length: 2582
Score: 137.00 Matches: 83
Percent Similarity: 34.90% Conservative: 58
Best Local Similarity: 20.54% Mismatches: 147
Query Match: 3.29% Indels: 116
DB: 11 Gaps: 14

US-09-765-271-56 (1-796) x AK016539 (1-2582)
QY 370 AspSerAsnSerSerLeuValSerGlnLeuValArgLysValGlyTyrValPhe 389
DB 1304 GAGAGAATCTGTCATTGGAGGAGAGATGCTGGCCAGGAGCAAGCATCAGATGGAG 1363
QY 390 GluGluLysGlyLeuSerArgTyrValPheAlaLysAspLeuProSerGluThrValLys 409
DB 1364 GAAGAGAAA-----AGACTGGCCCTGGAGAAAGACAGAGAGGCGCTGAAGAGGTGAAG 1414
QY 410 AsnLeuGluSerLysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLys 429
DB 1415 CATCTGCAGAGAGGAGGAGAGAACTGAGGGAGCAAGAAACAGAGACTGGCCAGGAGAAAG 1474
QY 430 GluAsnValAlaProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeu----- 446
DB 1475 GAGAGATAGCCAGAGAGAAAGACAGACTCAAGTAAAGAGAGAGATGGCCCAAAATG 1534
QY 447 -----LeuThrGlu 449
DB 1535 GAAGAAAGTCTGAGCAGGAGAGGACACAAGCTGTCTTTGGAAAAAGAAAGACTGACCGAA 1594
QY 450 AlaHisLysAlaLeuPhe***Asn-----LysGlyArgAsnSerAspPhe 464
DB 1595 AGGGAAAGGCTCTCCATGACAGAGAAAGAAAGCTCTTGAAGGGAAG---GAGCAATG 1651
QY 465 GlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn---LysGluLys 483
DB 1652 CAGGACATAAAGCGCAAACTGGAGACATACAGGAGAGAGCTGATTAACTTAAGTAAGAAG 1711
QY 484 LeuValAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArgLeuGlyLys 503
DB 1712 CTGATGAGGAGAGAGACTGTCTTTGTGGCG-----AAGAAGGAGAGCTGACTGTC 1762
QY 504 ProAsnSerGlnIleGluTyrThrGlu----- 512
DB 1763 ACAGAGAAGAAATAGATAGATGGAAGATATGCTTCTCGCGAGAGAGCAAGAAAGTTGGCT 1822
QY 513 ---AspGluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIle 531
DB 1823 GTGGAAAAAGTAAAGTTAGGAGAGAGATCAACGAGAGTGTCCAGATTAACCAACTTATC 1882
QY 532 PheAspGluHisAspIleLeuSerAspGluGlyAspAlaTyrValThrProHisMetGly 551
DB 1883 AGAGAAGAGACTGAGATTATTAAGAA----- 1909
QY 552 HisSerHisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAla 571
DB 1910 -----GCAGAGAGACTGACCTCCGAATGAAAGGTTGGCCAGGAG 1951
QY 572 TyrThrLysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnPro 591
DB 1952 AAGATGGAAGCTGGTGGAGAGATTGGAAGAGCCCTTAAGACAGAGATTAG----- 2002
QY 592 ThrGlyAspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeu 611
DB 2003 -----GAAGCCATCTTGGAGGAGGAGGCTGGCCCATGATGAGGTAGATCTA 2050

```

```

QY 612 ValArgLeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIle 631
DB 2051 ATGAGAAGAGAGAGTCTGATAGAGAGACATC----- 2083
QY 632 ProHisLysAspHisTyrHisAsnIleLysPheAlaTrpPheAspHisThrTyrLys 651
DB 2083 ----- 2083
QY 652 AlaProAsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHis 671
DB 2084 -----TCTGCAGAGAGAGCTGGCTGCTGACTCTGAAA----- 2113
QY 672 ProAspGluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGly 691
DB 2114 CCAGAGGAAGTGGCTGAAGAGAGAGGA-----AGAAGCAAAAAAGAAAGGA 2164
QY 692 LysLysAspHisSerGluAspProAsnLysAsnPheLysAlaAspGluProValGlu 711
DB 2165 ACATGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2215
QY 712 GluThrProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGlnLeu 731
DB 2216 GAGCAAAATGATTGAACCAAAA-----GAAGCAATAACAGAGCGAGCTTGAGAGAAAGCTG 2269
QY 732 LysGluAlaGlu 735
DB 2270 TCTTGGTAGAA 2281

RESULT 16
AY416214 2757 bp DNA linear GSS 12-DEC-2003
LOCUS Homo sapiens HCM5816 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY416214
VERSION AY416214.1 GI:39772174
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2757)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sainsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 2757)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sainsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..2757
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>2757
/locus_tag="HCM5816"

gene
Alignment Scores: 4.87 Length: 2757
Pred. No.: 137.00 Matches: 132
Score: 32.15% Percent Similarity: 32.15% Conservative: 95

```


Best Local Similarity: 18.70% Mismatches: 245
 Query Match: 3.29% Indels: 234
 DB: 29 Gaps: 31

US-09-765-271-56 (1-796) x AY416214 (1-2757)

127 ArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSerGlnHisArgGluGlyGly 146
 46 AGATCAGAGTCTAGTCTCGATCAGATCAGATCTCAATCTAGAAAGAGGATACAGT 105
 147 ThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGlnGlyArgThrThrAsp 166
 106 TCTAGGTCCTGTTCCAGAACATATTCAAGGTCCTGATGATGATGATGATGATGAT 165
 167 Asp-----GlyTyrIlePhe 171
 166 GATTATCGTCGCGATTACAGAAATAATAGAGAAATGAGACGACCTTATGGGTAC----- 219
 172 AsnAlaSerAspIleIleGluAspThrGlyAspAlaTyrIleValProHisGlyAspHis 191
 220 -----AGAGGAAGCGGTAGAGGTATTATCAAGGAGGAGGAGGTAGA 261
 192 TyrHisTyrIleProLysAsnGluLeuSerAlaSerGluLeuAlaAlaGluAlaPhe 211
 262 TATCAT----- 267
 212 LeuSerGlyArgGly-----AsnLeuSerAsnSerArgThrTyrArg 225
 268 -----CGAGTGGTTATAGACCTGCTCGAATAGAGGACCTCTAGAGTCTCTAGA 318
 226 ArgGlnAsnSerAspAsnThrSerArgThrArgThrValProSerValSerAsnProGly 245
 319 CGAGGTCGTCA-----CGTTCAGAGGTCCTCAAGAGATCGGTTCTCTCAAGA 372
 246 ThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAlaSerGlnSerAsn 265
 373 TCCAGAGCAGATCTCGCGGTCATATAGATCTTCTAGGTCCTCAAGATCATCTCTCT 432
 266 AspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSerGlnArgHisValGlu 285
 433 CGTTCCTCATCC-----CCATATAGCAAAATCTCTGTTCTTAAAGACGAGGCTCT 483
 286 SerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThrAlaArgGlyValAla 305
 484 CAGGAA-----AAACAAACCAAAAGCTGAAGG 513
 306 ValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMetSerGluLeuGluGlu 325
 514 GAACCCCAAGAGAGAGT-----CCGTTGAAAGTAAATACACAGGAGAACCG 561
 326 Arg-----IleAlaArgIleIleProLeuArg 334
 562 AAGANN 621
 335 TyrArgSerAsnHisTrp-----ValProAspSerArgProGluInPro--- 349
 622 NNNNNNNNNNNNTGGCGCTTTCAGCTTATGATAATAGTCTAGATACCCCAT 681
 350 SerProGlnPro-----ThrProGluProSerProGlyProGlnProAlaProAsn 366
 682 AGTCCTTCACCTATTGTGTACACACCTAGTCAGAGTTTCATCTTGTCTGATGCTCCCATG 741
 367 LeuLysIle-----AspSerAsnSerSerLeuValSerGln 378
 742 CTCAGTACAGTTCCTCTGCAAAAAATACTCTCTCAGCATTCATTCATTCATTCAGCAT 801
 379 LeuValArgValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrVal 398
 802 AGTCCTGAAAGGTCGGTCTGGTCTGTT-----GGAATGGATCTAGTCGATACAGT 855
 399 PheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGln 418
 856 CTTTCTCAGATAGTCCA-----ATTTCATCACATCCCTTCCAGGAAGGTCCTGCA 906

QY 419 GluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGlu 438
 DB 907 AAGCAATCGCA-----CCACAGAT---GCTCCAAGAGATGAG--- 942
 QY 439 PheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLys 458
 DB 943 -----TCTAGGGCCCGTTCTCTGCTTTTATCTCTGAT 972
 QY 459 GlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSer 478
 DB 973 GGTGGAGATCAGGAACATCGCAAGACGCTGGAGAGTCTTAAAGAGGTTCCACAGATCAAG 1032
 QY 479 Thr-----AsnLysGluLys 483
 DB 1033 TCTAGAGTATTCTCGCTGTGATAGGGTAATACAGGGGATAAGAGGCTTCAAAAGAGAAA 1092
 QY 484 LeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArgLeuGlyLys 503
 DB 1093 -----GGATCA 1098
 QY 504 ProAsnSerGlnIleGlu---TyrThrGluAspGluValArgIleAlaGlnLeuAlaAsp 522
 DB 1099 GAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1158
 QY 523 -----LysTyrThrThrSerAspGlyTyrIlePheAspGluHisAsp 536
 DB 1159 AAGAGTCTGGAACAAAGATTTAATGATTCAAGGGGATGACACAGAGGAGGAGGAGGAG 1218
 QY 537 -----IleIleSerAspGluGlyAspAlaTyrValThrPro 548
 DB 1219 GATTATAGACAGTTCAGAGAGTCTCTCGACATCAGGTAAGAGTTTGTCTACT--- 1275
 QY 549 HisMetGlyHisSerHisTrpIleGlyAspSerLeuSerAspLysGluLysValAla 568
 DB 1276 -----GCA 1278
 QY 569 AlaGlnAlaTyrThrLysGluLysGlyIleLeuProSerProAspAlaAspValLys 588
 DB 1279 TCTCACCAGGAACTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1335
 QY 589 AlaAsnProThrGlyAspSerAlaAlaIleTyrAsnArgValLysGlyGluLysArg 608
 DB 1336 GGCNATAGAGAAAGTGTGTA-----TTAGAGAAAGAAAGAAAT 1374
 QY 609 IleProLeuValArgLeuProTyrMetValGluHisThrValGluValLysAsnGlyAsn 628
 DB 1375 TATAAACTTAAAGAGAGTGTGATATGTAGTGGAAAGGCTAGCAGTACACAAA----- 1425
 QY 629 LeuIleIleProHisLysAspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHis 648
 DB 1426 -----GATAAGCAC 1434
 QY 649 ThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyr 668
 DB 1435 AAGAGAAAGACAAAATTTCTGAAGAGATA-----ACAGTAAAGAAAGAA 1479
 QY 669 ValGluHisProAspGluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHis 688
 DB 1480 ACTCAGTCACCTCAGCAG-----GTAAAGTCTGAAAG 1512
 QY 689 ValLeuGlyLysLysAspHisSerGluAspProAsnLysAsnPheLysAlaAspGluGlu 708
 DB 1513 CTCAAGACCTCTTTGATTATGATCAGTCCCTCTTACACAGAAATCTGGATGACGAGAAAG 1572
 QY 709 ProVal-----GluGluThrProAlaGluProGluValProGlnValThrGluLys 726
 DB 1573 TCTACCTTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1632
 QY 727 ValGluAlaGlnLeuLysGluAlaGluValLeuAlaLysValThrAspSerSerLeu 746
 DB 1633 CCTGAAGTCAAACTCAAAATGGCACCTGTTCTCTCTT-----GATGATTCTAACAGA 1683

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaehiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Okawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 4561)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Sushiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

A cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers
1. 4561
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM_DB:E230012G01"
/db_xref="MGI:2425940"
/db_xref="taxon:10090"
/clone="E230012G01"
/sex="female"
/tissue_type="oviduct"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="2 days pregnant adult"
337..4560
/note="laminin, alpha 4 (MGD|MGI:109321, GB|Y09827, evidence: BLASTN, 99%, match=4527)
putative"

misc_feature

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

ORIGIN

Alignment Scores:
Pred. No.: 15.2 Length: 4561
Score: 135.00 Matches: 140
Percent Similarity: 34.91% Conservative: 104
Best Local Similarity: 20.03% Mismatches: 257
Query Match: 3.24% Indels: 198
DB: 11 Gaps: 34

US-09-765-271-56 (1-796) x AK054014 (1-4561)

QY 8 GlnAlaArgThrValLysGlu-----AsnAsnArgValSerTyrIleAsp 22
Db 1474 CAAGCCACACATGAGGACAAATCAAGAGATCAACAGAGATGCTCTACTATGGA 1533

QY 23 GlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLysArgGlu 42
Db 1534 -----GAGAATCAGGAACCTTGGCCCGAGGAATAGCTGAGAAGCTG 1575

QY 43 GlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThrSerHis 62
Db 1576 ---GTGTTGGCCAGAAAATGCTGGAAGAGGTCCGAAGCCGCGAGCCATTCCTACCCAT 1632

QY 63 GlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSerGluGlu 82
Db 1633 CGGAG-----CTGTTGGATGAGGAGGAGGATGAGGCCGAGAA 1671

QY 83 LeuLeuMetLysAspProAsnTyrLysLeuLysAspGluLeuValAsnGluValLys 102
Db 1672 TTGCTGAGCAAGCTGAGAACTGG-----CAGCGCTGCACAATGACACCCGC 1719

QY 103 GlyGlyTyrValIleLysValAsp-----GlyLysTyrTyrValTyrLysAspAla 120
Db 1720 TCITTTATCCCGTGGTCTGGAGCAGCTGGACGACTACACGTAAGCTGTCCGACCTC 1779

QY 121 AlaHisAla-----AspAsnValArgThrLysGluGluIleAsnArg--- 134
Db 1780 CAGGAATCCATTAAACAGGCCCTCGACCATGTCTAGGATGTCAGAAAGACATGAATAGAGCC 1839

QY 135 -----GlnLysGlnGluHisSerGlnHisArgGluGlyGlyThrProArgAsn 150
Db 1840 ATCACTTTCAAGCAGCGGACCATGAGAAACAACATGAGAGTGAAGGAACACATGGAA 1899

QY 151 AspGlyAlaValAlaLeuAlaArgSerGlnGlyArgTyrThrThrAspAspGlyTyrIle 170
Db 1900 GTTGTGGTGCTCTCTGATGATGCTCTGACAGCTCTCTTACCATCCTCAGCTCACTCTT 1959

QY 171 PheAsnAlaSerAspIleLeuGluAspThrGlyAspAlaTyrIleValProHisGlyAsp 190
Db 1960 GAGAACTTGCATGATGATTAAGAATGATGATGATGATGATGATGATGATGATGATGATG 2016

QY 191 HisTyrHisTyrIleProLysAsnGluLeuSerAlaSerGluLeuAlaAlaGluAla 210
Db 2017 -----GCCAAAATGAGCTCAAGGAAA----- 2040

QY 211 PheLeuSerGlyArgGlyAsnLeuSerAsn-----SerArgThr 223
Db 2041 ---CTATCCAACTGAGTAACCTCAGTCATGCTTGGTTTCAGGAAGCTACGGACCATGCA 2097

QY 224 TyrArg---ArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSerValSer 242
Db 2098 TACAATCTTCAACAGGAGCCGATGAGCTAAGCAGAAATTTGACAGTTCAGACATGAAC 2157

QY 243 AsnProGlyThrThrAsnThrAsnThrSerAsn---AsnSerAsnThrAsnSerGlnAla 261
Db 2158 GGGCTGTCAGAAAGGCTTTGGATGTCATCAACGCTCTATGAAAATATCGCCAAATATGTC 2217

QY 262 SerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSerGln 281
Db 2218 AGTGAAGCCCAACGAAACAGCAGAACTT----- 2244

QY 282 ArgHisValGluSerGlyLeuValPheAspProAlaGlnIleThrSerArgThrAla 301

Db	2245	-----GCTCTGAATATCACTGATCGAATTTAT	2271
Qy	302	ArgGlyValAlaValPro---HisGlyAspHis-TyrHisPheIleProTyrSerGlnMe	320
Db	2272	GATGCTGTGAGTGGGAGTTGACACGACGATCATTTACCAT	2311
Qy	320	tSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTr	340
Db	2312	-----AAGGATGAAGTGACACCT	2331
Qy	340	pValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGl	360
Db	2332	TCTCAATCAAGCAGCAGAGCTGCAGGCCAAGCAGAT	2368
Qy	360	yProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVa	380
Db	2369	-----TCTAGCAATGATGAAGCAGTGGCTGCACACGAG	2400
Qy	380	LaGlyValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAl	400
Db	2401	CAGCGGTGGGT---GGAGCCCTGTGGAGGAAGGCGCCCTC	2440
Qy	400	aLysAspLeuProSerGluThrValLysAsnLeu	411
Db	2441	-AGAGCAGACTGAATCATCTCTTAAAGCACTACAGCGCAGCAGAGGGGACGCCCA	2499
Qy	412	-----GluSerLysLeuSerLysGlnGluSerValSerHisThrLeuThrAl	427
Db	2500	CCAGCGCTGGCGCAGTCCCAAGCTCTTCAATTGAGGAAGCT---AAACAAGACGACAGCGGC	2556
Qy	427	aLysLysGluAsnValAlaProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLe	447
Db	2557	TGTCACACAGGTWACACACATGCTGAACCTCAGCACTGGTCCAGACCTGCA	2616
Qy	447	uThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLe	467
Db	2617	GACC-----TTTGATCATCTGCATA	2637
Qy	467	uAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsnLysGluLysLeuValAspAs	487
Db	2638	TAACACTGCAGTGGACTCTCTCGGGACGAGTGAGAAACCTCAGCGAGTTGTCCGCCA	2697
Qy	487	pLeuLeuAlaPheAlaProIleThrHisProGluArgLeuGlyLysProAsnSerGl	507
Db	2698	GCCTCTGGATCAGCTGTGATCTGGAGCAGAG--CGCCCTGCAAGCAA	2745
Qy	507	nileGluTyrThrGluAspGluValArg-----IleAlaGln-----LeuAlaAs	522
Db	2746	CATTTCCGACAGATCCAGAGATCCGAGAGCTCATGTCTCAACACGAGGTGTCGAAG	2805
Qy	522	pLysTyrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAspGluGl	542
Db	2806	CAAGATCCAAGTCTCC-----ATGATGTTGAT-----GG	2835
Qy	542	yAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSerLeuSe	562
Db	2836	CCAGTCAGCTGTGGAAGTGAC-----CCCAAGTCAAGTGTGA	2874
Qy	562	rAsp---LysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuProPr	581
Db	2875	TGACTGAAGCCCTTACATCCATCAGCTGTGTATCATGAGCCT-----CCTCC	2922
Qy	581	oSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyrAs	601
Db	2923	AAAGCCG-----GCAGAGCCCACTGGGGCTGGGTAGCAGATCAGTTGT	2967
Qy	601	nArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGluHisTh	621
Db	2968	CCTCTACTCGGAGCAAAAAC-----GCCAAAAAGAAATACATG-----GG	3009
Qy	621	rValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle-L	641

```

Db      1346  ACACAAACAAATAATCNCACCTCAACGATACAGATACATATTT 1288
Qy      272  LysGlnLeuTyrLysLeuProLeuSerGlnArgHisValGluSerAspGlyLeuValPhe 291
Db      1287  CAGAAATAGAACAAACCAACGAAAC--CAATACAGGCACAA-GAAGAAATAGGAAAGTCATA 1232
Qy      292  AspProAlaGlnIleThrSerArgThrAlaArgGlyValAlaValProHisGlyAsp-Hi 311
Db      1231  GAACCG-----ACAACACAGACC-----ACGCCAACAAACCAACAA 1196
Qy      311  sTyrHisPheIleProTyrSerGlnMetSerGluLeuGluGluArgIleAlaArgIleI 331
Db      1195  CTACCCAAAGCACAAGCTCAAGACAGTCAACACCTTACCACACCCCAACCAACAAACATA 1136
Qy      331  eProLeuArgTyrArgSerHisTyrValProAspSerArgProGluGlnProSerPr 351
Db      1135  AAAAAAAAAAAAAAGCAAAACAAATAATATGCGAAGACATACATAAAAC-CCAAGC-- 1079
Qy      351  oGlnProThrProGluProSerProGlyProGlnProAlaProAsnLeuLysIleAspSe 371
Db      1078  -----ACACCAAGCGCTTACCACCAACCAACCAACCAACCAACATACATACATAFA----- 1031
Qy      371  rAsnSerSerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGl 391
Db      1030  AACACAAAGCGCATAGAAACAA----- 1010
Qy      391  uLysGlyIleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLe 411
Db      1010  ----- 1010
Qy      411  uGluSerLysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAs 431
Db      1009  -AAAAAAAAAAAAAAAAAAAAACAAACACACCCACCCCAACAAATAATAAAACCAACCC 951
Qy      431  nValAla-----ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuTh 448
Db      950  AACCGCACACACGAAACCAACAAACAAACAAACCAACCAACCAACCAACCAACCA 891
Qy      448  rGluAlaHisLys-AlaLeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeu 468
Db      890  CCACCCCAACAAACACGCAACAAACAAACAAACAAACAAACAAACAAACAAACAA 831
Qy      468  spLysLeuLeuGluArgLeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspL 488
Db      830  GGAACACAAACACACACACACACACACACACACACACAA----- 790
Qy      488  euLeuAlaPheLeuAlaProIleThrHisProGluArgLeuGlyLysProAsnSerGlnI 508
Db      789  -----AATCCACACGACCAAAATTACAGCCCAACCAACCAACAA 756
Qy      508  leGluTyrThrGluAspGluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerA 528
Db      755  GAAACAAACAGAGACAGACAAACA-----CAAGAACAAATAATATAACCAACACAG 706
Qy      528  spGlyTyrIlePheAspGluHisAspIleIleSerAspGluGlyAspAlaTyrValThrP 548
Db      705  AAGAGAGAACTTCAAGACCAAC-----ACAA 679
Qy      548  roHisMetGlyHisSerHisTyrIleGlyLysAspSerLeuSerAspLysGluLysVala 568
Db      678  GACACAAACCCACACAAACAC--ACAACAAATTAACCATACCAACAAACAACTCAACAA 622
Qy      568  laAlaGlnAlaTyrThrLysGluLys-GlyIleLeuProProSerProAspAlaAspVal 587
Db      621  ACACCAAAATGTATCATGAAGCCAAATAAATCTATCGAAAAAATACCTCAACAAACAA 562
Qy      588  LysAlaAsnProThrGlyAspSerAlaAlaIleTyr----- 600
Db      561  CCACCAACACGACGACCCACGATCTCAACAACTACCAACAAACAAACAAACAAAGAA 502
Qy      601  ---AsnArgValLysGlyGluLysArgIleProLeuVal-----ArgLeuProTyr 616

```

```

Db      501  CACCACCGCTAAAAAATAAAAAAAGAAATACCCATGTATCTCTCATATGTCCCCAC 442
Qy      617  Met-----ValGluHisThrVal---GluValLysAsnGlyAsnLeuIlePro 632
Db      441  CAACAACCCCAATACACACCAATATAGAAATGTTTAATCTCAAGATACCAATACCA 382
Qy      633  HisLysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAla 652
Db      381  CACAATGAC-----CAACAACATACAA-----AAACACACACACCCCTCTG 343
Qy      653  ProAsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisPro 672
Db      342  CCCCCCACCACACATAAAAA-----AAACACCG 310
Qy      673  AspGluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLys 692
Db      309  GACAAA-----ACACACACACCCACACACAAA 283
Qy      693  LysAspHisSerGluAspProAsnLysAsnPheLysAlaAspGluGluProValGluGlu 712
Db      282  CAACAACATCAACATCCCATCAACAAACGAAAGTACCAAGAAATACACCCACACACA 223
Qy      713  ThrProAlaGluProGluValProGlnValGluThrGluLys 726
Db      222  CAACCC---AAACCAAAACAAACAAACCAACCAACCAACAA 184

```

RESULT 21

CG756662 1652 bp DNA linear GSS 24-OCT-2003
 LOCUS P051-4-D06.ya Ppa EcoRI BAC Library Pristionchus pacificus genomic,
 genomic survey sequence.

ACCESSION CG756662
 VERSION CG756662.1 GI:37984450

KEYWORDS GSS.

SOURCE Pristionchus pacificus

ORGANISM Pristionchus pacificus

REFERENCE 1 (bases 1 to 1652)

AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,

Buntjer,J., van der Meulen,M. and Sommer,R.J.

TITLE An integrated physical and genetic map of the nematode Pristionchus

pacificus

JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)

MEDLINE 22835951

PUBMED 12884007

COMMENT Contact: Sommer RJ

Evolutionary Biology

Max-Planck-Institute for Developmental Biology

Spemannstr. 37-39, Tuebingen D-72076, Germany

tel: 00497071601371

Fax: 00497071601498

Email: ralf.sommer@tuebingen.mpg.de

Class: BAC ends.

FEATURES

Location/Qualifiers

1..1652

/organism="Pristionchus pacificus"

/mol_type="genomic DNA"

/strain="California"

/db_xref="taxon:54126"

/clone_lib="Ppa EcoRI BAC Library"

/note="The library was generated by a partial digest of

the genomic DNA with EcoRI and cloning into the BAC

vector."

ORIGIN

Alignment Scores:

Pred. No.: 3.59 Length: 1652

Score: 134.50 Matches: 116

Percent Similarity: 30.17% Conservative: 65

Best Local Similarity: 19.33% Mismatches: 218

Query Match: 3.23% Indels: 203

DB: 29 Gaps: 24

Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.

FEATURES

```
1. 1755
location/Qualifiers
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by
the genomic DNA with EcoRI and clo
test="
```

ORIGIN

Alignment Scores:	
Pred. No.:	3.94
Score:	134.50
Percent Similarity:	33.46%
Best Local Similarity:	20.66%
Query Match:	3.23%
DB:	29
	Gaps: 19
	Indels: 143
	Mismatches: 238
	Conservative: 70
	Matches: 113
	Length: 1755

US-09-765-271-56 (1-796) x CG754413 (1-1755)

[illegible]

D	995	CAC	TATATAAGAAACAACAAGCAAAATCA-----AGT
Q	443	Ala	TyrAsnLeuLeuThrGluAla-HisLysAlaLeuPhe***AsnLysGlyArgAsnSe
D	965	TC	AATCATCTTAACACTTACTCAATCATTAATCAATCAATCAATCAATCAATCAAC--
Q	462	rAsp	PheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsnLysGl
D	907	--	---AATTAACCCGGCTACAATAAGCA885
Q	482	wLys	LeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHis-----497
D	884	CCA	ATTCGCCACCACCTCAGCACCCCTCTTCACCCCTCAACAACACACACCCACCCACCA
Q	498	-Pro	GluArgLeuGlyLysProAsnSerGlnIleGlyTyrThrGluAspGlu-----514
D	824	ACA	CCCCAACACTCACACACCCCCTCACTCGAACAAACAAACAAAACCCACACC765
Q	515	----	ValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAs533
D	764	CAC	CACCTCAATACCTCCCTCAACCGGCACACACCAAGCCACACCCCAACAATCTTCAA705
Q	533	pGlu	HisAspIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSe553
D	704	AAC	CACACACATA-----AAGCAACACCATCACACCCACCA705
Q	553	rHis	TrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrTh573
D	665	TC	CAAAACTA-----GCAAGTTACACCAAC642
Q	573	rLys	GluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGl593
D	641	TACT	CACCGAGACAACCCCACTCATCGGAAGAATAATACTGTAAACACACA--584
Q	593	yAsp	SerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValAr613
D	583	----	-----CACATAACTCGATCTC-----566
Q	613	gLeu	ProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHi633
D	565	-----	TACCACATCCACACACATCAAAATAAAAAACAAACACTCC-CTCTTCCA514
Q	633	sLys	AspHis-----TyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyLy651
D	513	CAC	CTCCACAGAATATACAAACATCACTCAACACTATACATTAATTAACGCATACATATCA454
Q	651	sala-	-----ProAsnGlyTyrThrLeuGluAs660
D	453	CATA	ACCTACTCCATCANTTGACATACCCAACTACATCCCTTACACATACACACCAACCA394
Q	660	pLeu	PheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAs680
D	393	AC	CAATACACGCCACACCCACCCACACACCGCCGACACACGT-----348
Q	680	pGly	TrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAsp-----698
D	347	-----	-----AAACCAAAACACACCCCAACATATT322
Q	699	-----	ProAsnLysAsnPheLysAlaAspGluProValGluThrProAl715
D	321	ACA	CAATACCAACCAACCAAGCAAAAAAAAAAACAACAAACAAACACACACTCCAC262
Q	715	aGlu	Pro---GluValPro720
D	261	CAC	ACCTTACAAAAATACCA243

RESULT 23

RE30U1 23
AY411030

HLH Locus

DEFINITION

2
3
4
5
6
7
8
9
10
11
12

ACCESSION

VERSION

4773 bp DNA linear GSS 16-DEC-2003
, VIRTUAL TRANSCRIPT, partial sequence,

```

KEYWORDS  GSS.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 4773)
AUTHORS    Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,
            Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
            Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
            Adams, M.D. and Cargill, M.
TITLE      Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
JOURNAL    Science 302 (5652), 1960-1963 (2003)
PUBMED     14671302
REFERENCE  2 (bases 1 to 4773)
AUTHORS    Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,
            Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
            Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
            Adams, M.D. and Cargill, M.
TITLE      Direct Submission
JOURNAL    Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT    This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES   Location/Qualifiers
            source          1..4773
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        <!.>4773
                        /gene="RBBP6"
                        /locus_tag="HCM4088"
ORIGIN
Alignment Scores:
Pred. No.:      17.9      Length:      4773
Score:          134.50    Matches:    137
Percent Similarity: 34.10%  Conservative: 131
Best Local Similarity: 17.43%  Mismatches: 304
Query Match:     3.23%      Indels:     214
DB:              29        Gaps:        28

US-09-765-271-56 (1-796) x AY411030 (1-4773)
QY      24  LysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLysArgGluGly 43
DB      2581 AAGAGGAGACTCCGAGACTGACATATCAATCAATCATCTCTCTCAGAGGATGAA 2640
QY      44  IleAsnAlaGluGlnIleValIleLysIleThrAsp-----GlnGlyTyrValThr 60
DB      2641 -----AAATCACTGGAAACCCCGAGAAAGCTCACTCTAAA 2676
QY      61  SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
DB      2677 TCAGCAAAAGAACACCAAGAA-----ACAAACCAAGTCAAA 2712
QY      81  GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
DB      2713 GAGGAAAGTGAAGAGACTATTCGAAGATGTCAATCAGAAAGCTCACTAAG 2772
QY      101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
DB      2773 -----GAAGAAAG 2781
QY      121 AlaHisAlaAspAsnValArgThrLysGluIleAsnArgGlnLysGlnGluHisSer 140
DB      2782 GCCAAGAACCTAATGAGAAACAAACCACTTGATAATAGGAGAGAGAGAGAGAGAGAG 2841
QY      141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyValAlaValAlaLeuAlaArgSerGln 160
DB      2842 AAAACTGAAGAAAGCGCTAGATAAGATTTTGAGCTCTCTCAATGAAATCTCGAA 2901
QY      161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180

```

```

2902 -----CTAGAAGTGACTGAAATAGTGAACCATCA 2931
181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIlePro---LysAsnGlu 199
2932 CCAAAAGCGCAAAATGGAACCTGATCTAGTAAAAAATGGATAGACCCCTGAAAGGACAAA 2991
200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
2992 ATTTCCTTAAAGT-----CGCCAGCCCAAAAATCAAACTCAACAGA----- 3033
220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239
3033 ----- 3033
240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
3034 -----GAAACTGGGAAGAAATTTGGAAGTACAGAAAATATATCAACACCAAGAA 3084
260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGln---LeuTyrLysLeuPro 278
3085 CCCTCTGAAAAATTTGAGTCAACATCTAGCAAGTTAAACAGAAAAAGTCAAGGAAAG 3144
279 LeuSerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSer 298
3145 GTCAGACGAAAGTCTGCTGGAACCTCAAGGATCCAGCTCAACTCTGTGTGATTACACAGT 3204
299 ArgThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSer 318
3205 ACGAGCTCAACTCGAGGCGAGT-----CCTGTGCGG 3234
319 GlnMetSerGluLeuGluGluAlaArgIleAlaArgIleIleProLeuArgTyrArgSerAsn 338
3235 AAATCTGAGAAAAACAGATACAAAGCCAACTGTGATTAAACAGTGAAGAAATATAT 3294
339 His-----TrpValProAspSerArg 345
3295 AATGACAATATCCGCGCCAGCTGAAGATGTTATCATTTATGATTCAGGTTCTCTCAATCCAAA 3354
346 ProGluGlnProSerProGlnProThrProGluProSerProGlyProGlnProAlaPro 365
3355 TGGGATAAAGATGACTTTGAAATCTGAAGAAAGATGTTAAATCCACAGCCTATATCA 3414
366 AsnLeuLysIleAspSerAsnSerLeuValSerGlnLeuValArgLysValGlyGlu 385
3415 AGT-----GTAGGAAAACTCTAGTGTATATAAAAAATGTTAGTACAAAG----- 3459
386 GlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAlaLysAspLeuProSer 405
3460 -----CCATCAAAATATAGTCAAGTATCTCTGAGAAAAGAAAGTAGGCGCATCC 3504
406 GluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSerValSerHisThrLeu 425
3505 GAGAAAAATTCAGAAATTCACCAAGGACGTGAGC---CATGAAATCATACAACTAGGTT 3561
426 ThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyrAspLysAlaTyrAsn 445
3562 AAAAGTTCAAAAAATCTGCTCATCTAGTGAAGGAGGAAACCAAGATCGAGATTATTC 3621
446 LeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArgAsnSerAspPheGln 465
3622 GTGTGGAAAAAGAGAACCCCT-----GAAAAGAGGAGAAACAGCACTCAGCCA 3669
466 AlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsnLysGluLysLeuVal 485
3670 GAGAAAGAGAGTAAATTTGAGCCGTCTGAATGACACAGGAAATTTTAAAGT----- 3720
486 AspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArgLeuGlyLysProAsn 505
3720 ----- 3720
506 SerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeuAlaAspLysTyrThr 525

```

3721	Db	----	CTGTCTCAATCTTCCAAAGAGCGGTAGA	3747
526	QY	ThSer	AspGlyTyrIlePheAspGluHisAspIleIleSerAspGluGlyAspAlaTyr	545
3748	Db	ACGTCA	-----GATAACATGATCCACTCGTCTTCTCCTCAATAAAGAC	3792
546	QY	ValThrProHisMetGlyHisSerHisThrIleGlyLysAspSerLeuSerAspLysGlu	565	
3793	Db	TTCACTCCCAATAGA	-----GACAAAAAACTGACTATGACACACAGA	3834
566	QY	LysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuProSerProAspAla	585	
3835	Db	GAGTATTCAGTTCCAAACGTAGAGTGAAGAAGATTAACAAGACGAAGAAGACTCT	3894	
586	QY	AspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyrAsnArgValLysGly	605	
3895	Db	CCTTCTCGG	-----AATAAAGATTCTGTCATCGACAGAAAAATAAACCAAGGAA	3945
606	QY	GluLysArgIleProLeuValArgLeuProTyrMetValGluHisThrValGluValLys	625	
3946	Db	GAGAGAGATTGGCT	-----AAAAAGGAACAGAGAGATTCCAA	3984
626	QY	AsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIleLysPheAlaTyrPhe	645	
3985	Db	AAAAAGTAAT	---TCTAGTCCCTCAAGAGACAGA-----	4014
646	QY	AspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAlaThrIle	665	
4015	Db	AAACCTCATGATCAAAAGCC	-----	4035
666	QY	LysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp	683	
4036	Db	ACTTATGATACTAAACGGCCAAATGAAGACAGACAAAATCTGTAGTAAAAATCCTTGTAA	4095	
684	QY	AsnAlaSerGluHisValLeuGlyLysAspHisSerGluAspProAsnLysAsnPhe	703	
4096	Db	GATCGTGAGAAGCATGTATTAGAACGAAGCAACAATAAAGAGTCAAGTGCAATAAACA	4155	
704	QY	LysAlaAspGluProValGluGluThrProAlaGluProGluValProGluValGlu	723	
4156	Db	CTTTATATCTTAACCCACAGAG	-----ACACAGGTTGAA	4191
724	QY	ThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeu	742	
4192	Db	AAAGACAAATTACTGGCAAAATTGACAGAGTACTGTCAAGCTAAACCCAGTTAAGT	4251	
743	QY	AspSerSer	---LeuLysAlaAsnAlaThr-GluThrLeuAlaGlyLeuArgAsnAsnLe	761
4252	Db	CATTCTCTAGACTTTCTCTGACTTAACTAGAGAACTG	-----	4291
761	QY	uThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLe	781	
4292	Db	----ATGAAGCTGCTTTTGACACAGACTATATGAAGGTGACAGTGAAGATTAATGTTCT	4347	
781	QY	uLeuLysGlySerAsn	786	
4348	Db	GTAAGAAGAGAGGAAT	4363	

RESULT 24	CG755974	2041 bp	DNA	linear	GSS 24-OCT-2003
LOCUS	CG755974/c				
DEFINITION	P051-2-E12.Ya Ppa EcoRI BAC Library Pristionchus pacificus genomic, genomic survey sequence.				
ACCESSION	CG755974				
VERSION	CG755974.1	GI:37983025			
KEYWORDS	GSS.				
SOURCE	Pristionchus pacificus				
ORGANISM	Pristionchus pacificus				
	Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.				
REFERENCE	1 (bases 1 to 2041)				
AUTHORS	Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer R.J.				

1648	ACAACCAACACACACCCCAACACACACACCAACCAACCCACCCACCTTCTACCCCCCGCCGAC	1589
361	----ProGlnProAlaProAsnLeuIysIleAspSerAsnSerSerLeuValSerGlnLe	379
1588	ATAGACCAACCAACCAACGAC	1567
379	uValArgIysValGlyGluGlyTyrValPheGluGluGlyGlyIleSerArgTyrValPh	399
1566	----ACACACCAACCAAAACAACTCTCTACAGCGGTGCA	1532
399	eAlaLysAspLeuProSerGluThrValIysAsnLeuGluSerIysLeuSerIys	417
1531	CAGCCCCACATGCTCCCAACATAAAAAC--AAAAAAACACAAAAATAAA	1475
418	----GlnGluSerValSerHisThrLeuThrAla	427
1474	AACATAACAGACCATCATCCCCACACACAAACCGCCCCAACGCCACCCACCGCCAC	1415
427	----	427
1414	ACNTAGAACACCGCAGACATCTACGACACACCCANAGCCGACCTTCTACNACCACA	1355
428	----LysIysGluAsnVal	434
1354	AACCAACCAACAAAAAAACATAAAATATCAAGCCACACAGCAACCAACAC	1295
434	oArgAspGlnGluPheTyrAspIysAlaTyrAsnLeuLeuThrGluAlaHisIysAlaLe	454
1294	TCACACAAAGCAAACTATCGCGNCGCG--ACCCAGACAGCAACAAAT	1247
454	uPhe***AsnIysGlyArgAsnSerAspPheGlnAlaLeuAspIysLeuLeuGluArgLe	474
1246	ATACAAACCAACAAACACACACACACACACGCAACATGCAACCAACAAAAACACACAA	1187
474	uAsnAspGluSerThrAsnIysGluIysLeuValAspAspLeuLeuAlaPheLeuAlaPr	494
1186	CCAGACCCACAAACACAGCAACCAACCAAAACANAGACCAACA--CACACACA	1136
494	oIleThrHisProGluArgLeuGlyIysProAsnSerGlnIleGluTyrThrGluAspG	514
1135	AACMACACAACAGACACCAATACACAGACACACACACACCAACCAACCAACTAGCA	1076
514	uValArgIleAlaGlnLeuAlaAspIysTyrThrThrSerAsp--GlyTyrIlePh	532
1075	CACACATACCAACTCTACACGCGACACACACACACACAAATACGCGCACACACACA	1016
532	eAspGluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHi	552
1015	CAAAACACACACACACAGCACACACACCCCAACACACACACACACACACACAGCA	956
552	sSerHisTrpIleGlyIysAspSerLeuSerAspIysGluIysValAlaGlnAlaTy	572
955	ACAGCATACGTACTCTACCACCACTCTCCAGAAC--	922
572	rThrLysGluIysGlyIleLeuProProSerProAspAlaAspValIysAlaAsnProTh	592
921	-----CCAACAAACGAAACCAATCATATAAACACACACAAC	887
592	rGlyAspSerAlaAlaIleTyrAsnArgValIysGlyGluIysArgIleProLeuVa	612
886	ACA--CAACACAGTCTCCAGCCCAACCAACCATACAG--	846
612	lArgLeuProTyrMetValGluHisThrValGluValIysAsnGlyAsnLeuIlePr	632
845	AAGAAACGCAAAATATCACACACACAAACACCAACCCACCCACTCATCAACACACAC	786
632	oHisIysAspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLys--	651
785	AAGAACGACATATAAACACCAACAAA-----CAAAACACGGAAACACACACACA	732
652	----AlaProAsnGlyTyrThrLeuGluAspLeuPheAlaThr--	664
731	CAGCACACACAGACGCGGGGAGACATAAACACACACACACACACCAACACGCAAAAAA	672

Qy	665		-----IleIysTyrTyValGluHisProAs	673
			:::	
Db	671	CACAAAGACCACCCACACAAAAA	CAATCACCACACACACCCCACA	612
Qy	673	pGluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLy	693	
		:::	:::	
Db	611	CCACCGCACACACCCCAAAACACAATAAGAACAACAACACAC-	-	566
Qy	693	sAspHisSerGluaspPro-	-	699
		:::	:::	
Db	565	-CCGCACACACCAAACCCCCCTGAACCTCCAACACCAACATAACATGAAAAG	507	
Qy	700	AsnLYsAsnPheLYsAlaAspGluGluProValGluGluThr-	-	713
		:::	:::	
Db	506	AACGAAAAACAAACACCGGTCAACGGTAACCAACTCACACACAAACACACACCATCCA	447	
Qy	714	-ProAlaGluProGluValProGlnValGluThrGluLYsValGluAlaGlnLeuLYsgl	733	
		:::	:::	
Db	446	ACCACCAACCCCAACCAACACACACCAAAAAAAAAAAAAACAAACCAACCAAGGGA	387	
Qy	733	uAlaGluValLeuLeuAlaLYsValThrAspSerSerLeuLYsAlaAsn	749	
		::: :::	::: :::	
Db	386	GAGACAA-CTCAAAAAACACACACACCCCAACCAACCAAAATCTCAT	344	
RESULT 25				
AY404466				
LOCUS				
DEFINITION Homo sapiens HCM1900 gene, VIRUAL TRANSCRIPT, partial sequence,				
AY404466 linear GSS 12-DEC-2000				
3714 bp DNA				

RESULT 25	AY404466	3714 bp	DNA	linear	GSS 12-DEC-2003
LOCUS	AY404466				
DEFINITION	Homo sapiens HCM1900 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.				
ACCESSION	AY404466				
VERSION	AY404466.1	GI:39760443			
KEYWORDS	GSS.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 3714) Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.				
TITLE	Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios				
JOURNAL	Science 302 (5652), 1960-1963 (2003)				
PUBMED	14671302				
AUTHORS	2. (bases 1 to 3714) Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA				
COMMENT	This sequence was made by sequencing genomic exons and ordering them based on alignment.				
FEATURES	source				
	1..3714				
	Location/Qualifiers				
	/organism="Homo sapiens"				
	/mol_type="genomic DNA"				
gene	/db_xref="taxon:9606"				
	<1..>3714				
	/locus_tag="HCM1900"				
ORIGIN					
Alignment Scores:					
Pred. No.:	13.4	Length:	3714		
Score:	134.00	Matches:	183		
Percent Similarity:	34.60%	Conservative:	135		
Best Local Similarity:	19.91%	Mismatches:	350		
Query Match:	3.22%	Indels:	254		
DR:	29	Gaps:	41		

Alignment Scores:		
Fred. No.:	13.4	Length: 3714
Score:	134.00	Matches: 183
Percent Similarity:	34.60%	Conservative: 135
Best Local Similarity:	19.91%	Mismatches: 350
Query Match:	3.22%	Indels: 254
DR:	29	Gaps: 41

US-09-765-271-56 (1-796) x AY404466 (1-3714)

```

Qy 31 GluAsnLeuThrProAspGluValSerIysArgGluGlyIleAsnAlaGluGlnIleVal 50
Db 7 GAGAGGATACCAAGAGAGAGGTGCCCAAGGAAGATGGAGTT-----48
Qy 51 IleLysIleThrAspGlnGlyTyThrSerHisGlyAspHisTyThrHisTyThrAsn 70
Db 49 -----GGTGATGCAACATTTTCGATTCC 72
Qy 71 GlyLysValProTyrAsp---AlaIleIleSerGluGluLeuLeuMetLysAspProAsn 89
Db 73 AGCAAGATTATGGAAGAGCATAGTACATACAAAAGGAGCTAGATTTACAA-----AAT 126
Qy 90 TyrLysLeuLysAspGluAspIleValAsnGluValLysGlyGlyTyThrValIleLysVal 109
Db 127 GGTAGCTTAGGAAGAGACTCTGTGGTGCATTCCTGTGAGAACGATTCCCAAAACATGATG 186
Qy 110 AspGly-----LysTyThrValTyThrLysAspAlaAlaHisAlaAspAsn 125
Db 187 GAGAGCCTCAGCCCAAGAAATATCTTCCAGTCTGAGATTAAAGCCAATGGAGACTAT 246
Qy 126 ValArgThrLysGluGluIleAsnArgGln---LysGlnGluHisSerGlnHisArgGlu 144
Db 247 TCTGGCTCTCTATTAAACCTCTCAACACCTGTGCTGCAAGAGAGCCCTTCTCCTTTG 306
Qy 145 GlyGlyThrProArgAsnAspGlyValAlaValAlaLeuAlaArgSerGlnGly---ArgTyr 163
Db 307 GGAACAGTGTCAAGAC-----AGCCCTCTCTAGCCAAATCCAGGGAAGCAAGCAG 360
Qy 164 ThrTyrAspGlyTyThrIlePheAsn-----AlaSerAspIleIleGlu 178
Db 361 TTCTTTATGATGGAAGTGCACAAAATATTCCTATGAAACCTCCAACCTCTTACTCAAC 420
Qy 179 AspThrGly-----AspAlaTyThrIleValProHisGlyAspHisTyThrIlePro 196
Db 421 ACTACATCTCCCTCAGTGGATATCCACTTGAAGAGAGCAGACTTTGATCATATTAAGTGC 480
Qy 197 LysAsnGluLeuSerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGly 216
Db 481 CGGGAC-----AGTGAAGGGCCCTG-----501
Qy 217 AsnLeuSerAsnSerArgThrTyArgArgGlnAsnSerAspAsnThrSerArgThrAsn 236
Db 502 AGGCTCTCAGAGAGCCTCCCTATTCCAATATAGCTCAAGGCATAATCGCATGACAAAT 561
Qy 237 -----Tyr 237
Db 562 GTCTACTCTCTTGAGGGCTGGAAGGTGGAAGGCATCGCTCGCTCCTGGCCATGTGG 621
Qy 238 -----ValProSerValSerAsnProGlyThrThrAsnThrAsnThrSerAsn 253
Db 622 AATGGAAGTTCCCTTGATGTATGTGGCCGCTCTATATCAGCAGATCGGAGCGCGAAGC 681
Qy 254 AsnSerAsnThrAsnSerGlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGln 273
Db 682 ATGCCTTCAGCCCAAGCAAGCAGCAAAATAGCATTAGCAGACACCTTGGCGCTTCAA 741
Qy 274 LeuTyThrLysLeuProLeuSerGlnArgHisValGluSerAspGlyLeuValPheAspPro 293
Db 742 -----CCCAAGTTAACTAGACACAAAGGAG-----CTTGCATCTGAAAC 780
Qy 294 AlaGlnIleThrSerArgThrAlaArgGlyValAlaValProHisGlyAspHisTyThrHis 313
Db 781 ATCAATTGAGAACTAGGAAGTACTCCAGCAGCAGCCTGAGTCACATGGGAGCGCTAC---837
Qy 314 PheIleProTyrSerGlnMetSerGluLeuGluArgIleAlaArgIleIleProLeu 333
Db 838 -----AGCCCATCATTCCCAAG 855
Qy 334 ArgTyrArgSer-----AsnHisTrpValProAspSerArgProGluGlnProSerPro 351
Db 856 TTGTACAGAGCCACAGAGAACCCAGCTGACACCTCTCAGCTTGCCTTCCCAAGAAATCTCTCTG 915

```

```

Qy 352 GlnProThrProGluProSerProGlyProGlnProAlaProAsnLeuLysIleAsp---370
Db 916 GGCATTTCCAAACGACAAACTTTGGGAAAGGATCTACCTCATAGCTAGTAATAGACAAT 975
Qy 371 -----SerAsnSerSerLeuValSerGlnLeuValArg 381
Db 976 GACATTTACCTTAATTTTCTTTGAGCTCAGGGCTTTACCCCTATAAAACCTCTGCT 1035
Qy 382 LysValGlyGluGlyTyThrValPheGluGluLysGlyIle-----SerArgTyThrValPhe 399
Db 1036 TCTGAAGGCAATCCTTAATGTAAGTTTACCTCAGTGTCCCTGCGCAGTCCAGGAGTGGCT 1095
Qy 400 AlalysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGlu 419
Db 1096 CGGAAGATGCTTCTGGCTCTCCACCTCTCTGCTGCTCTGATGACTTTTGATCAGGCTTCA 1155
Qy 420 SerVal-----SerHisThrLeuThr-----Ala 427
Db 1156 TATGTGGGACAAACCCGAGTCAITCACTTCTGCTGAGAGTGCAGACAGATTTTTCGG 1215
Qy 428 LysLysGluAsnValAlaProArgAspGlnGluPheTyThrAspLysAlaTyThrAsnLeuLeu 447
Db 1216 ACCAGGAGGAATCTCTCTGTGATCTGTGAAATTTGATGAGGAGATTTGGAAGCCTC 1275
Qy 448 ThrGluAla-----HisLysAlaLeuPhe***AsnLysGlyArgAsnSerAsp 463
Db 1276 AGACGCTCCTCAGAACCCCGACCTGCTTTCGGGAACGGAACGATATTAGTCTC 1335
Qy 464 PheGlnAlaLeuAspLysLeuLeu-----GluArgLeuAsn 475
Db 1336 ATTTGAGGAGCTGATGACCTTGATGATTATCACCGCGGAGAGGAGGAGAAAGTCTCAGG 1395
Qy 476 AspGluSerThrAsnLys-----GluLysLeuValAspAspLeuLeuAlaPheLeuAla 493
Db 1396 GAGCAGAAATGGAGCGATTTGGAGACAGCGTCTGAGAGCACTCTCAGTCTCTGTGCT 1455
Qy 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGlu-----Tyr 510
Db 1456 GAATACACA-----AAGCCTGACAGTCTGCTTATCTACTGAGGACACC 1497
Qy 511 ThrGluAspGluValArgIleAlaGlnAlaAspLysTyThrThrSerAspGlyTyr 530
Db 1498 GTGAAGATGTGAGAAATCAACAGGAGCTTGAGAGCTTGAGAGTCTCTGATGAGGAG 1557
Qy 531 IlePheAspGluHisAspIleIleSerAspGluGlyAspAlaTyThrValThrProHisMet 550
Db 1558 TCTGTGTTGAGGAAGCCCTCATGAGCCCTGAC-----ACAAGATACAGG 1602
Qy 551 GlyHisSerHisTrpIleGlyLysAspSerLysSerAspLysGluLysValAlaAlaGln 570
Db 1603 TGCACCGG-----AAAGACTCCTCCTGATGACAGACTTGGCAAGCTGTGGG 1650
Qy 571 AlaTyThrLysGluLysGly-----IleLeuProProSerProAspAlaAsp---Val 587
Db 1651 AGTCTAGTCAGAGCAGTGCACGCTTTTACCCCGAGGAGCACCAGGAATGATGAATA 1710
Qy 588 LysAlaAsnProThrGlyAspSerAlaAlaIleTyThrAsnArgValLysGlyLys 607
Db 1711 CTCAGTGA-CCTCAGCCGAGTCTCTCCACCATCTCTCCACCTTTCGAAAGCTTCCAG 1769
Qy 608 ArgIleProLeuValArgLeuProTyrMetValGluHisThrValGluValLysAsnGly 627
Db 1770 CGAGT-CCTCTTATTAAGTATCTCTACCAAGACCCAGAGGGGTATAAGTGAAGAAGACA 1828
Qy 628 AsnLeuIleIleProHisLysAspHisTyThrHisAsnIleLysPheAlaTyThrPheAspAsp 647
Db 1829 GATCTC-----AGGAGTGTGCTGCAAA---1849
Qy 648 HisThrTyThrLysAlaProAsnGlyTyr-----ThrLeuGluAspLeuPheAlaThr 664
Db 1850 -----TGGAAAGAAACCCGAGATA-GTCATTTCTGAACAACTCTCAGGAACTTAAGCAAAA 1902

```

```

Qy 665 IleLysTyrTyrValGluHisProAspGluArg----- 675
Db 1903 ATCAAAGACATAAATGATCAGATGATGATGCTTTTCAGAGAGTTGGATGGAATGTGCT 1962
Qy 675 ----- 675
Db 1963 CTTTGGATGGAGACAGAAATCTGAAACAACCTAATGAACTTATGAAGGAGAAGAGATTG 2022
Qy 676 ProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHis 695
Db 2023 GATCATCTAAACCGGAAAATAGCTGAACCTGGAAAACATTTGTTGGTGAAGAACCAAG 2082
Qy 696 SerGluAsp-----ProAsnLysAsnPheLysAlaAspGluGluProValGlu 711
Db 2083 GATGCTGACCTGTTGGATGTTCAAGACCAACACTTTGAA----- 2121
Qy 712 GluThrProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGlnLeu 731
Db 2122 -----GACCTGGAGTTCCAGCAGCTTGAACATGAGAGCCGCTAGATGAAGAA 2169
Qy 732 LysGlu-----AlaGluValLeuLeuAlaLysValThrAsp----- 743
Db 2170 AAGGAGAACTTGACTCAACAGCTCTGCTGGAAGTTGCTGAATATCAACGGAAACATCGTT 2229
Qy 744 -----SerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeu 757
Db 2230 TCTAGAAAGGAAAAATTTCGCATTGAAAAAGCAAGCCAAATCACATTTGTCAGCAGGCT 2289
Qy 758 -----ArgAsnAsnLeuThrLeuGlnIleMetAsp 767
Db 2290 CAGAGAGACAGATCATTTTGTGAAGAAAAGAAATAATTTAATAATGATGTTGCAAGA 2349
Qy 768 AsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer----- 785
Db 2350 GAAAAGGAGAACTTTTGTAAATTTGGAAAAGAAATACTCCAGCCTCTCTGGGGGGAAGGG 2409
Qy 786 -----AsnProSerSer-----ValSerLysGluLysIleAsn 796
Db 2410 TTTCCCGTTAAACCCCAATACTTTAAAGAGGGCTATATCAGTGTAAATGAGATTAT 2466

```

Search completed: October 1, 2004, 11:30:12
Job time : 5074 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 1, 2004, 07:39:37 ; Search time 139 Seconds

(without alignments)
3177.992 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 4165

Sequence: 1 SYELGLYQARTVKNRVSY.....KLIALIKGNPSSVSKEKIN 796

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool.h/US09765271/runat 30092004 113754 12303/app query.fasta_1.967
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=90 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=25
-MODE=LOCAL -OUTPMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09765271 @cgn2_1_69 @runat 30092004 113754 12303 -NCPUS=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4163	100.0	2389	3	US-08-961-083-55
2	4163	100.0	2389	4	US-09-536-784-55
3	4163	100.0	2451	4	US-09-468-656A-9
4	4163	100.0	8195	4	US-08-961-527-94
5	2786.5	66.9	2531	4	US-09-468-656A-11
6	2772	66.6	2531	4	US-09-468-656A-5
7	2649.5	63.6	2290	3	US-08-961-083-65
8	2649.5	63.6	2290	4	US-09-536-784-55
9	1774	42.6	2359	4	US-08-961-527-192
10	1222.5	29.4	6867	4	US-08-961-527-192
11	1203.5	28.9	1455	4	US-09-468-656A-7
12	1187.5	28.5	1342	3	US-08-961-083-181

13	1187.5	28.5	1342	4	US-09-536-784-181
14	988	23.7	973	4	US-08-961-527-355
15	770.5	18.5	1684	4	US-08-961-527-258
16	304.5	7.3	841	2	US-08-743-637B-34
17	304.5	7.3	841	3	US-08-526-840B-34
18	170	4.1	8155	4	US-08-956-171E-63
19	169	4.1	12173	4	US-08-956-171E-310
20	163	3.9	11091	4	US-09-134-001C-2243
21	159	3.8	30549	4	US-09-134-001C-322
22	152.5	3.7	31096	4	US-08-956-171E-59
23	150.5	3.6	2616	4	US-09-417-197-108
24	148	3.6	14736	4	US-08-961-527-171
25	147.5	3.5	2745	4	US-08-817-514F-5
26	147.5	3.5	15213	4	US-08-961-527-26
27	147	3.5	2439	4	US-09-513-783A-5
28	146	3.5	3095	6	5231168-1
29	146	3.5	5400	4	US-09-134-000C-1773
30	145.5	3.5	4223	3	US-09-541-782-5
31	145.5	3.5	4223	4	US-09-723-820-5
32	145.5	3.5	4223	4	US-10-270-085-5
33	144	3.5	3121	3	US-08-961-083-117
34	144	3.5	3121	4	US-09-536-784-117
35	144	3.5	14066	4	US-09-601-198-56
36	143	3.4	4522	4	US-08-956-171E-39
37	143	3.4	8100	4	US-09-554-337-4
38	143	3.4	11517	1	US-07-920-281C-1
39	143	3.4	11517	3	US-08-466-277-1
40	143	3.4	15538	4	US-09-554-337-1
41	141.5	3.4	3945	4	US-09-200-650B-6
42	141.5	3.4	7941	4	US-09-816-703A-1
43	139.5	3.3	4948	4	US-09-562-702A-23
44	139.5	3.3	4948	4	US-09-561-818A-23
45	139.5	3.3	4972	4	US-09-562-702A-27
46	139.5	3.3	5306	4	US-09-562-702A-21
47	139.5	3.3	5306	4	US-09-561-818A-21
48	139.5	3.3	5330	3	US-09-562-702A-25
49	138	3.3	6744	1	US-08-119-125A-2
50	138	3.3	8982	3	US-08-976-255-5
51	137.5	3.3	4833	4	US-09-513-783A-21
52	137	3.3	6948	4	US-09-543-681A-1262
53	136.5	3.3	3827	2	US-08-447-031A-1
54	136	3.3	2439	4	US-09-513-783A-3
55	136	3.3	3171	4	US-09-513-783A-31
56	136	3.3	3279	5	PCT-US93-03077-2
57	135.5	3.3	4643	4	US-09-453-702B-22
58	135.5	3.3	4899	2	US-08-210-535-5
59	135	3.2	2582	2	US-08-816-105A-2
60	135	3.2	2592	1	US-09-604-958-3
61	135	3.2	5009	1	US-08-487-890A-3
62	135	3.2	5009	2	US-08-478-435-3
63	135	3.2	5009	2	US-08-337-483-3
64	135	3.2	5009	2	US-08-478-373-3
65	135	3.2	5009	3	US-08-474-671-3
66	135	3.2	5009	3	US-08-483-577A-3
67	135	3.2	5009	3	US-08-897-438-3
68	135	3.2	5009	4	US-08-637-654-3
69	135	3.2	5009	4	US-08-649-518-3
70	135	3.2	5033	1	US-08-487-890A-2
71	135	3.2	5033	2	US-08-478-435-2
72	135	3.2	5033	2	US-08-337-483-2
73	135	3.2	5033	2	US-08-478-373-2
74	135	3.2	5033	3	US-08-474-671-2
75	135	3.2	5033	3	US-08-483-577A-2
76	135	3.2	5033	3	US-08-897-438-2
77	135	3.2	5033	4	US-08-637-654-2
78	135	3.2	5033	4	US-08-649-518-2
79	134.5	3.2	4899	6	5268270-1
80	134.5	3.2	5194	1	US-09-599-652-1
81	134.5	3.2	5194	2	US-08-642-846-1
82	134.5	3.2	5194	4	US-09-264-604-1
83	134.5	3.2	5194	4	US-10-002-389-1
84	134.5	3.2	590073	4	US-08-545-5280-1
85	134	3.2	6924	1	US-08-015-973-2

Sequence 181, Appl
Sequence 355, Appl
Sequence 258, Appl
Sequence 34, Appl
Sequence 34, Appl
Sequence 63, Appl
Sequence 310, Appl
Sequence 2443, Appl
Sequence 322, Appl
Sequence 59, Appl
Sequence 108, Appl
Sequence 171, Appl
Sequence 5, Appl
Sequence 26, Appl
Sequence 5, Appl
Patent No. 5231168
Sequence 1773, Appl
Sequence 5, Appl
Sequence 5, Appl
Sequence 5, Appl
Sequence 117, Appl
Sequence 117, Appl
Sequence 56, Appl
Sequence 39, Appl
Sequence 4, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 6, Appl
Sequence 1, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 21, Appl
Sequence 21, Appl
Sequence 21, Appl
Sequence 25, Appl
Sequence 2, Appl
Sequence 5, Appl
Sequence 21, Appl
Sequence 1262, Appl
Sequence 1, Appl
Sequence 3, Appl
Sequence 31, Appl
Sequence 2, Appl
Sequence 22, Appl
Sequence 5, Appl
Sequence 2, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Patent No. 5268270
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 2, Appl

86 134 3.2 6924 2 US-08-448-164-2
 87 134 3.2 9626 4 US-09-150-867-2
 88 133.5 3.2 4612 2 US-08-447-031A-8
 89 132 3.2 2817 4 US-09-620-312D-1085
 90 132 3.2 4797 4 US-09-976-594-988

ALIGNMENTS

RESULT 1

US-08-961-083-55
 ; Sequence 55, Application US/08961083
 ; Patent No. 6159469
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/961,083
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 55:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2389 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; US-08-961-083-55

Alignment Scores:
 Pred. No.: 0 Length: 2389
 Score: 4163.00 Matches: 796
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.95% Indels: 0
 DB: 3 Gaps: 0

US-09-765-271-56 (1-796) x US-08-961-083-55 (1-2389)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnArgValSerTyr 20
 DB 2 TCTTACGAGTGGGACTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTCTTCCAT 61
 QY 21 IleAspGlyLeuGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
 DB 62 ATAGATGGAACCAACGCCAGCAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCAAG 121
 QY 41 ArgGluGlyLeuAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
 DB 122 CGTGAAGGAATCAATCTGAGCAATTCGTATCAAGATTAACAGACCAAGGCTATGTCAC 181

QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
 DB 182 TCACATGCGGACCCTATCATTTACAAATGGTAAGGTTCTTATGAGCGCTATCATCAGT 241
 QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
 DB 242 GAAGAATTACTCATGAAAGATCCAAACTATAAGCTTAAAGATGAGGATATTGTTAATGAG 301
 QY 101 ValLysGlyGlyTyrValLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
 DB 302 GTCAAGGCTGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACTTAAGGATGCT 361
 QY 121 AlaHisAlaAspAsnValArgThrLysGluGluLeuAsnArgGlnLysGlnHisSer 140
 DB 362 GCCCAGCGGGATACGTCCTGACAAAGAGGAATCAATCGACAAAAACAAGAGCATAGT 421
 QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaAArgSerGln 160
 DB 422 CAACATCGTGAAGGTGGAACCTCCAAGAAACGATGGTGTGCTTGCCTTCGACGTTCCGAA 481
 QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
 DB 482 GGACGCTATATCTACAGATGATGGTTATATCTTTAANGCTTCTGATATCATAGAGATACT 541
 QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
 DB 542 GGTGATGCTTATATCGTTCCTCATGGAGATCATTTACCATTACATTTCTTAAGAATGAGTTA 601
 QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
 DB 602 TCAGCTAGCAGTTCGGCTGCTGACAGAGCCTTCTTATCTGTGTCGAGGAAATCTGTCAAT 661
 QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
 DB 662 TCAAGAACCTATCGCCGACAAATAGCGATTAACACTTCAAGAACAACTGGGTACTCTTCT 721
 QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
 DB 722 GTAAGCAATCCAGGAACCTACAATACTAACACAAACAGCAACACAGCAACAGTCAA 781
 QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
 DB 782 GCAAGTCAAGTAATGACATTTGATAGTCTCTTGAACAGCTCTACAAACTGCTTTGAGT 841
 QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
 DB 842 CAACGACATGTAGAAATCTGATGGCCCTTGTCTTTGATCCAGCAACAATCACAAGTCAACA 901
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
 DB 902 GCTAGAGGTGTTGACGTGCCACAGGAGATCATTTACCACCTTCATCCCTTACTCTCAATG 961
 QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340
 DB 962 TCTGAATTTGGAAGAACGAAATCGCTGATTTATTCCTCTGCTTATTCGTTTCAACCATGG 1021
 QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
 DB 1022 GTACCAGATTTCAGGCCAGAACCAAGTCCACAAACCGACTCCGGAACCTAGTCCAGGC 1081
 QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
 DB 1082 CGCAACCTGACACCAATCTTAAAAATAGACTCAAAATCTCTCTTGGTTAGTCAGCTGTA 1141
 QY 381 ArgLysValGlyGlyTyrValPheGluGlyLysGlyIleSerArgTyrValPheAla 400
 DB 1142 CGAAAGTTGGGAGGATGATGTTTCGAAAGAAAGGGCACTCTCTGTTATGCTTTGGC 1201
 QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
 DB 1202 AAAGATTACCATCTGAAACTGTTAAAAATCTTTGAAGCAAGTATCAAAACAAGAGAGT 1261
 QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440


```
Db 1262 GTTTCACACACTTTAACTGCTAAAAAGAAAAATGTTGCTCTCGTGACCAAGATTTTAT 1321
Qy 441 AsplysialatyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db 1322 GATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATAAAGGTCGT 1381
Qy 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
Db 1382 AATTCTGATTTCCAGCCTTAGACAAATATTAGAACGCTTGAATGATGAATCCACTAAT 1441
Qy 481 LysGluLysLeuValAspAspLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db 1442 AAAGAAAAATTTGGTAGATGATTTATTTGGCAATTCCTAGCACCAATACCATCCAGAGCGA 1501
Qy 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
Db 1502 CTTGGCAAAACCAAAATTCCTCAATTTGAGTATATCTGAAGACGAAGTTCGTATTGCTCAATTA 1561
Qy 521 AlaAspLysTyrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
Db 1562 GCTGATAGATATACAACTGATGCTTACATTTTGTGATGAACATGATATTAATCAGTGAT 1621
Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560
Db 1622 GAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1681
Qy 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
Db 1682 CTTTCTCATAGGAAAAAGTTGACGCTCAAGCCTATATCAAGAAAAAGGATGATCCTACCT 1741
Qy 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600
Db 1742 CCATCTCCAGCGAGATGTTAAAGCAATCCAACTGGAGATGTCAGACGCTATTATTC 1801
Qy 601 AsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
Db 1802 AATCGTGTAAAGGGGAAAAACGAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1861
Qy 621 ThrValGluValLysAsnGlyAsnLeuIleLeuProHisLysAspHisTyrHisAsnIle 640
Db 1862 ACAGTTCAGGTTAAAAACGGTAATTTGATTTATTCCTCATAGGATCATTAACCATATATT 1921
Qy 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660
Db 1922 AATTTCTGTTGTTGATGATCACACATACAAAGCTCAAATGGCTATACCTTGGAAAGAT 1981
Qy 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
Db 1982 TTGTTTGGCAGCATTAAGTACTAGCTAGAACACCCCTGACGACGCTCCACATTTCTAATGAT 2041
Qy 681 GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
Db 2042 GGATGGGGCAATGCCAGTGAGCTGTTGTTAGGCAAGAAAGACCACTGAGATCCAAAT 2101
Qy 701 LysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720
Db 2102 AAGNACTTCAAGCGGATGAAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCCT 2161
Qy 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
Db 2162 CAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAAAGCAGAAAGTTTGTCTTGGCAAA 2221
Qy 741 ValThrAspSerSerLysLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
Db 2222 GTAACGATTTCTAGTCTGAAAGCCAAATGCAACAGAAACTCTAGCTGGTTTACGAAATAAT 2281
Qy 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780
Db 2282 TTGACTCTTCAATTTGGATTAACATAGTATCATGGCAGAGCAGAAAAATTTACTTGGC 2341
Qy 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796
```

Db 2342 TTGTTAAAAAGGAGTAATCTTTCATCTGTAGTAAGGAAAAATAAAC 2389

RESULT 2

US-09-536-784-55
; Sequence 55, Application US/09536784
; Patent No. 6573082

GENERAL INFORMATION:

; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/536,784

; FILING DATE: 30-Oct-1997

; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/961,083

; FILING DATE: OCT-30-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Michelle S. Marks

; REGISTRATION NUMBER: 41,971

; REFERENCE/DOCKET NUMBER: PB340P3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 55:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2389 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 55:

US-09-536-784-55

Alignment Scores:

Pred. No.: 0 Length: 2389

Score: 4163.00 Matches: 796

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 99.95% Indels: 0

DB: 4 Gaps: 0

US-09-765-271-56 (1-796) x US-09-536-784-55 (1-2389)

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20

Db 2 TCTTACAGTGGAGCTGTTATCAAGCTAGAACCGTTAAGGAAAAATAATCGTGTTCCTAT 61

Qy 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40

Db 62 ATAGATGGAANAACAGCGACGCAAAAACGAGAAATTTGACTCTCTGATGAGGTAGCAAG 121

Qy 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60

Db 122 CGTGAAGGAATCAATGCTGAGCAATTCGTATCAAGATAACAGACCAAGCTATGTCAC 181

Qy 61 SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer 80

Db 182 TCACATGGCGACCACTATCATATTATTAAATGGTAAAGGTTCCCTTATGACGCTATCAT 241

Qy 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100

Db 242 GAAGAAATTTACTCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATTTGTTAATGAG 301
Qy 101 VallysGlyGlyTyrValIleIysValAspGlyIysTyrTyrValTyrIleuIysAspAla 120
Db 302 CTCAGGGGTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAAGATGCT 361
Qy 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnLysHisSer 140
Db 362 GCCACGCGGATACGTCGTACAAAGAGGAATCAATCGCAAAACAAGAGCATAGT 421
Qy 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
Db 422 CAACATCGTGAAGGTGAACCTCAAGAACAGATGCTGCTGTGCTTGGCCTTGGCAGTTCGCAA 481
Qy 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db 482 GGACGCTATACACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGTACT 541
Qy 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db 542 GGTGATGCTTATATCGTTCCTCATGGAGATCATACCATTACATTCCTAAGATGAGTTA 601
Qy 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db 602 TCAGCTAGCGAGTTGGCTGCTGCAGAGCCTTCTATCTGTCGAGGAAATCTGTCAAAAT 661
Qy 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
Db 662 TCAAGAACCTATCTGGCGGCAAAATAGCGATAACATTCAGAACCAAACTGGGTACTCTTCT 721
Qy 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db 722 GTAAGCAATCCAGGAATACAAATACTAACACAGCAACAACAGCAACACTAACAGTCAA 781
Qy 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrIleLysLeuProLeuSer 280
Db 782 GCAAGTCAAAAGTAATGACATTTGATAGTCTCTTGAACACAGCTCTCAAACTGGCTTTGAGT 841
Qy 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db 842 CAACGACATGTAAATCTGATGCGCTGTCTTTGATCCAGCAACAATCAAACTCGAACCA 901
Qy 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db 902 GCTAGAGGTGTTGCAAGTCCACAGAGATCATACCCTTCATCCCTTACTCTCAAAATG 961
Qy 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340
Db 962 TCTGAATTTGGAAGAACGAATCGCTCGTATATTTCCTCTTCTGTTAGTCAGCTGTA 1021
Qy 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db 1022 GTACAGATTCAGGCCAGAACCAACCACTCCACACCACTCCGGAACTAGTCCAGGC 1081
Qy 361 ProGlnProAlaProAsnLeuIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
Db 1082 CCGCAACCTGCACCAATCTTAAATATAGACTCAATTCCTTCTGTTAGTCAGCTGTA 1141
Qy 381 ArgLysValGlyGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
Db 1142 CGAAAGTTTGGGAAGGATATGTTATTCGAAGAAAAGGGCATCTCTGTTATGCTTTGGC 1201
Qy 401 LysAspLeuProSerGluThrValLysAsnLeuLeuSerLysLeuSerLysGlnLeuSer 420
Db 1202 AAAGATTACCATCTGAACTGTAAATCTTGAAGCAAGTATCAAAACAAGAGACT 1261
Qy 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
Db 1262 GTTTCACACACTTAATCTGTAAGAAAAGAAATGTTGCTCTCTGTCGACCAAGAAATTTAT 1321
Qy 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db 1322 GATAAAGCATATAATCTGTTAACTGAGGCTCAAAAGCCTTGTTCGNAATAGGGTCTGT 1381

Qy 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
Db 1382 AATCTCTGATTTCCAAAGCCTTAGACAAATTTATAGAACGCTTGATGATGATCGACTAAT 1441
Qy 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db 1442 AAAGAAAAATTTGGTAGATGATTTATTTGGCATTTCTTAGCAACCAATTTACCCATCCAGAGCA 1501
Qy 501 LeuGlyLysProAsnSerGluIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
Db 1502 CTTGCAACCAAAATTTCTCAAAATTGAGTATCTGAAGACGAAGTTCTGATTTCTCAATTA 1561
Qy 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
Db 1562 GCTGATAGTATACAACTGATGTTTACATTTTGTATGAACATGATATATCATGAT 1621
Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560
Db 1622 GAAGGAGATGCAATATGTAACGCTCATATGGGCCATAGTCACTGATTTGGAAGATAGC 1681
Qy 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
Db 1682 CTTTCTGATAAGGAAAAAGTTGCGACTCAAGCCTATATCTAAAGAAAAAGGTATCTTACT 1741
Qy 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600
Db 1742 CCATCTCCAGAGCGCAGATGTTAAAGCAAAATCCCACTGGAGATAGTGCAGCAGCTATTTAC 1801
Qy 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
Db 1802 AATCGTGTGAAGGGGAAAAACGAATTCCTACTCGTTCGACTTCCATATATGTTGAGCAT 1861
Qy 621 ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle 640
Db 1862 ACAGTTGAGGTTAAACCGTAAATTTGATTTCTCTCATAGGATCATTTACCATATATTT 1921
Qy 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660
Db 1922 AATTTTGTCTGTTTGTATGATCATACACATACAAAGCTCCAAATGGCTATACCTTGAAGAT 1981
Qy 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
Db 1982 TTGTTTGGCAGCATTAAGTACTACGTAGAACAACCTCGAACACGTCACATTTCTAATGAT 2041
Qy 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
Db 2042 GGATGGGCAATCCCGATGAGCATGTTGAGCAAGAAAGACCAAGTGAAGATCCAAAT 2101
Qy 701 LysAsnPheLysAlaAspGluProValGluGluThrProAlaGluProGluValPro 720
Db 2102 AAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCCT 2161
Qy 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
Db 2162 CAAGTAGAGACTGAAAAAGTAGAACCCCACTCAAGAGACGAGAGTTTGTCTTGGAAA 2221
Qy 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsn 760
Db 2222 GTAACGATTTCTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGTTTACGAATAAT 2281
Qy 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780
Db 2282 TTGACTCTTCAAAATTTATGGATAACATAGTATCATGGCAGAGACGAGAAAAATTTACTTGG 2341
Qy 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796
Db 2342 TTGTTAAAGGAAGTAATCTTCTATCTGTAAGTAAAGAAAAATTAAC 2389

RESULT 3

US-09-468-656A-9

; Sequence 9, Application US/09468656A

; Patent No. 6582706

```

; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; CURRENT FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 2451
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)-(2451)
; OTHER INFORMATION: n = a, c, t or g
US-09-468-656A-9

Alignment Scores:
Pred. No.: 0 Length: 2451
Score: 4163.00 Matches: 796
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.95% Indels: 0
DB: 4 Gaps: 0

US-09-765-271-56 (1-796) x US-09-468-656A-9 (1-2451)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db 61 TCTTACAGTGGGACTGTATCAAGCTAGAACCGTAAAGGAAATAATCGTGTTCCTAT 120

QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 121 ATAGATGGAAACNAGCCAGCAGCAAAACCGAAGATTGTGCTCTGATGAGGTAGCAAG 180

QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db 181 CGTGAAGGAATCAATGCTGAGCAATCGTCATCAAGATAACAGACCAAGCTATGTCAC 240

QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db 241 TCACATGGCCGACCACTATCATATTACAAATGGTAAGGTTCCTTATGACGCTATCATCAG 300

QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db 301 GAAGAATTACTCATGAAGATCCAACTATAAGCTAAAGATGAGGATATTGTTAATGAG 360

QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db 361 GTCAAGGGTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGGATGCT 420

QY 121 AlaHisAlaAspAsnValArgThrLysGluLysLeuAsnArgGlnLysGlnGluHisSer 140
Db 421 GCCCAGCGGATACGTCGCTACAAAGAGGAATCAATCAAGAAACCAAGAGCATAGT 480

QY 141 GlnHisArgGluGlyThrProArgAspGlyValAlaValAlaLeuAlaArgSerGln 160
Db 481 CAACATCGTGAAGGTGAATCCNAGAAACGATGGTCTGTGCTTGGCAGCGTTCGCAA 540

QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db 541 GGACGCTATACATACAGATGATGGTTATATCTTTAATGCTCTGATATCATAGAGGATACT 600

QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db 601 GGTGATGCTTATATCGTTCCTCATGGAGATCATACCAATTACATTCCTCAAGAATGAGTTA 660

```

```

201 SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
661 TCAGCTAGCGAGTTGGCTGCTGCAGAAAGCCTTCTCTATCTGGTCAGGAAATCTGTCAAA 720
221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
721 TCAGAACCTTATCGCCACCAAAATAGCGATAACACTTCAGAACAAACTGGGTACTTCT 780
241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
781 GTAAGCAATCCAGGAATCTACAAATACTAACACAAAGCAACACAGCAACACTAACAGTCAA 840
261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
841 CCAAGTCAAAAGTAAATGATGATGATGCTCTTGAACACAGCTCTACAAACTGCCCTTGTAG 900
281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
901 CAACGACATGTAGAATCTGATGGCTTGTCTTTGATCCAGCACAAATCAACAGTCAACAA 960
301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
961 GCTAGAGGTGTTCAGTGCACACACGAGATCATACCACTTCATCCCTTACTCTCAAAATG 1020
321 SerGluLeuGluAlaArgIleAlaAqlleIleProLeuArgTyrArgSerAsnHisTrp 340
1021 TCTGAATTTGAAGAACCAATCGCTCGTATTATTCCTCTCGTATCTGTTTCAACCATTTGG 1080
341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
1081 GTACCAAGTTCAGGCCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGGC 1140
361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
1141 CCGCAACTGCACCAATCTTAAATATAGACTCAAAATCTCTTTGGTTAGTCAAGTGGTA 1200
381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
1201 CGAAAGTTGGGAAGGATATGATTATTCGAAGAAAGGAGCATCTCTCTGTTATGCTTTTGGC 1260
401 LysAspLeuProSerGluThrValLysAsnLeuSerLysLeuSerLysGlnGluSer 420
1261 AAGATTATCACTGAACTGTTTAAATCTTGAAGCAAGTTTCAAAACCAAGAGAGT 1320
421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
1321 GTTTCACACACTTAACTGCTAAAAAGAAATGTTGCTCTCTGTCGACCAAGAAATTTAT 1380
441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
1381 GATAAGCATATTAATCTGTTAACTGAGGCTCATAAAGCCTTGTITGNAATAAGGGTCGT 1440
461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
1441 AATTCTGATTTCAAGCCTTAGACAAATATTAGAAGCTTTGAATGATGATGACTAAT 1500
481 LysGluLysLeuValAspAspLeuAlaPheLeuAlaProIleThrHisProGluArg 500
1501 AAAAGAAAATTTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCGA 1560
501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValAlaGlnLeu 520
1561 CTTGGCAACCAAAATTTCTCAAATGAGTATCTGAAGACGAGTTCGTATGCTCAATTA 1620
521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
1621 GCTGATAAGTATACACAGCTCAGATGTTACATTTTGTATGATAACATGATAAATCAGTGAT 1680
541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560
1681 GAAGGAGATGCATATGTAAACGCTCATATGGGCCATAGTCACCTGGATTTGAAAGATAGC 1740
561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580

```

1741	Db	CTTTCTCATAGGAAGTTCAGCTCAAGCCCTATATAAGAAAAAGGTATCCTACCT	1800
581	Qy	ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaLalleTyr	600
1801	Db	CCATCTCCAGACCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTCGACGAGCTATTAC	1860
601	Qy	AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis	620
1861	Db	AATCGTGTGAAGGGGAAAAACGAATTCCTCGCTCCATATATGGTTGAGCAT	1920
621	Qy	ThrValGluValLysAsnGlyAsnLeuLelleProHisLysAspHisTyrHisAsnIle	640
1921	Db	ACAGTTGAGGTTAAAAACGGTAATTTGATTATTCTCTCATAGGATCATTACCATAATATT	1980
641	Qy	LysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp	660
1981	Db	AAATTTGCTTGGTTTGATGATCAACATACAGCTCCAAATGGCTATACCTTGGAAAGAT	2040
661	Qy	LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp	680
2041	Db	TTGTTTTCGCACGATTAAGTACTACGTAGAACACCCCTGACCAAGCTCCACATTTCTAATGAT	2100
681	Qy	GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysAspHisSerGluAspProAsn	700
2101	Db	GGATGGGGCAATGCGCAGTGAGCATGTGTGTAGGCAAGAAAGACCACACAGTGAAGATCCAAAT	2160
701	Qy	LysAsnPhelysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro	720
2161	Db	AAGAACTTCAAGCGGATGAAGACCCAGTAGAGGAANAACACTGCTGAGCCAGAGTCCCT	2220
721	Qy	GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys	740
2221	Db	CACTGAGACTGTAAAAAAGTAGAAGCCCACTCAAGAAAGCAGAAAGTTTGTCTGCAAA	2280
741	Qy	ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn	760
2281	Db	GTAACGGATTCCTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACCGAAATAAT	2340
761	Qy	LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla	780
2341	Db	TTGACTCTTCAAAATATGGATPAACAATAGTATCATGGCAGACGAGAAAATTACTTTGGC	2400
781	Qy	LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn	796
2401	Db	TTGTTAAAGCAAGTAACTCTTCATCTGTAGTAAGCAAAAAATAAAC	2448

RESULT 4

```

RES001 4
US-08-961-527-94
; Sequence 94, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

```


1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnArgValSer 19
 61 TCCTATGAGTTGGAGTTACCAAGCTGTCAGGATAAGAAAGAGTCTAATCGAGTTGCT 120
 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
 121 TATATAGATGGTCATCAGCTCGTCAAAAGGAGGAGAACTTGACACCAAGATGAAGTCAGT 180
 40 LysArgGluGlyLysAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
 181 AAGAGGAGGGGATCAACCCCAAAATGTTATCAAGATTACGGATCAAGGTTATG 240
 60 ThrSerHisGlyAspHisTyrHisTyrValAsnGlyLysValProTyrAspAlaIleIle 79
 241 ACCTCTCATGGAGACCAATATCATATATATATGCAAGGTTCCCTATGATGCCATCATC 300
 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLysLeuLysAspGluAspIleValAsn 99
 301 AGTGAAGAGCTCCTCATGAAGATCCGAATATATCATGTTGAAGGATTTCAGACATTTGTCAT 360
 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrValTyrLysLysAsp 119
 361 GAATCAAGGGTGGTATGTCATTAAGGTAACCGTAATACTATGTTTACCTTAAGGAT 420
 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
 421 GCRGCTCATGCGGATAATATTCGACAAAAGAGAGATTAAACGTCAGAGCAGGAACGC 480
 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
 481 AGTCATAATCATAACTCA-----AGAGCAGATAATGCTGTTGTCAGCCAGAGCC 531
 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
 532 CAAGGACGTTATACACCGATGATGGTATATCTTCAATGCACTCATATCAATGAGGAC 591
 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
 592 ACGGGTGATGCTTATGCTTCTCAGCGCAGCAATACCATTCATTCCTAAGAATGAG 651
 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
 652 TTATCAGCTAGGAGTTAGTCTCGAGAGGCTTATGGAATGGG----- 696
 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239
 697 -----AAGCAGGATCTCGCTCTCTTCAAGTCTAGTTATATGCA 738
 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
 739 AATCCAGCTCAACCAAGATTGTCAGAGAACCAATCTGACTGTCTCACTCCAATTTATCAT 798
 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
 799 CAA---AATCAAGGGGAAACATTTCAAGCCTTTACGTGAATTTGATGCTTAACCCCTTA 855
 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
 856 TCAGAACGCCATGTGGAATCTGATGGCTTATTTTCGACCCAGCGCAATCAAGTCGA 915
 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
 916 ACCGCGCAGAGGTGATGCTCCTCATGTGAACCATTACCACTTATCCCTTATGAACAA 975
 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
 976 ATGCTGAATTGAAAGAAAGAAATGCTGTTATTTCCCTTCCTGTTATGTTCAACACCAT 1035
 340 TrpValProAspSerArgProGlnProGlnProSerProGlnProThrProGluProSerPro 359
 1036 TGGGTACCAAGATTCAAGACCAAGAAACCAAGTCCACACCACTCCAGAACCTAGTCCA 1095
 360 GlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeu 379

1096 AGTCCGCAACCAAGCTCCCAAGCAATCAATTGAT-----GGCAATTTGCTCAAGAAGCT 1149
 380 ValArgLysValGlyGluGlyTyrValPheGluGluLysGlyLysSerArgTyrValPhe 399
 1150 GTTCGAAAGTAGGCGATGGTTATCTCTTTAGGAGAAATGGAGTTCTCGTTATATCCCA 1209
 400 AlaLysAspLeuProSerGluThrValLysAsnLeuLeuGluSerLysLeuSerLysGlnGlu 419
 1210 GCCAAGGATCTTTAGCAGAAACACAGCAGCGCATTTAGTAACTATGCGCAAGCAGGA 1269
 420 SerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPhe 439
 1270 AGTTTATCTCATAGCTAGGAACCTAAGAAACCTAGCACTCCCATCTAGTGATCGAATTT 1329
 440 TyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGly 459
 1330 TACAATAAGGCTTATGACTTACTAGCAAGAATTCACCAAGATTACTTGTATATAAAGT 1389
 460 ArgAsnSerAspPheGlnAlaLeuAspLysLeuGluAlaGluAsnAspGluSerThr 479
 1390 CGACAAGTTGATTTGAGCTTTGGATAACCTTTGGACGACTCAAGGATGCTCAAGT 1449
 480 AsnLysGluLysLeuValAspLeuAlaPheLeuAlaProIleThrHisProGlu 499
 1450 GATAAAGTCAAGTTAGTGAAGATATCTTGCTTCTTAGCTCCGATCGTCATCCAGAA 1509
 500 ArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGln 519
 1510 CGTTAGGAAACCAAAATGCGCAATTTACCTACCTGATGATGAGATCAAGTAGCCAAG 1569
 520 LeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSer 539
 1570 TTGGCAGCAAGTACACAGCAAGCGGTATATCTTTGATCTCTCGTATATAACAGT 1629
 540 AspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAsp 559
 1630 GATGAGGGGATGCTTATGTAACCTACATATGACCCATGACCTAGGATTAAGAAAGAT 1689
 560 SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeu 579
 1690 AGTTTGTCTGAAGCTGAGAGCGCGCAGCCAGGCTTATGCTTAAGAAAGAGTTTGACC 1749
 580 ProProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIle 599
 1750 CCTCTTCGACAGACCATCAGGATTCAGGAATACTGAGGCAAAAGGAGCAGAGCTATC 1809
 600 TyrAsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGlu 619
 1810 TACAACCGMTGAAAGCAGCTAAGAAAGGTGCCACTTGATCGTATGCTTACATCTTCAA 1869
 620 HisThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsn 639
 1870 TATACCTAGAGTCAAAACCGTATGTTAATCATACCTCATATTGACCATTTACCATAC 1929
 640 IleLysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGlu 659
 1930 ATCAATTTGAGTGGTTTGACGAGGCGCTTTATGAGGCACCTAAGGGGTATATCTTGAG 1989
 660 AspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsn 679
 1990 GATCTTTTGGGACCTGTCAAGTACTATGTGCAACATCCAAACGAAACGATCCGATTCAG 2049
 680 AspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspPro 699
 2050 AATGTTTTGTTAAGCTAGCGACCATGTTTCAAGAAACAAATAATGTTGCTCAAGCTATAC 2109
 700 AsnLysAsnPheLysAlaAspGluGlu-----ProValGluGluThr 713
 2110 AATCAACGGAACCAAGCAGGAGGAAACCTCAGACAGAAAAACCTGAGGAGAAAC 2169
 713 ----- 713

Db 2170 CCTCGAGAGAGAAACCGAAACGAGAGAACACAGAGTCTCCAAACCAACAGAGAGAACCA 2229
 Qy 714 -----ProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGln 730
 Db 2230 GAAGAAATCACAGAGAGATCAGAGAAACCTCAGGTGAGAGCTGAGAAAGGTTGAAGAAAAA 2289
 Qy 731 LeuLysGluAlaGluValLeuAlaLysValThrAspSerSerLeuLysAlaAsnAla 750
 Db 2290 CTGAGAGAGGCTGAGATTTACTTGGAAATATCCAGATCAATATCAAGTCCATGCC 2349
 Qy 751 ThrGluThrLeuAlaGlyLeuArgAsnLeuThrLeuGlnLeuMetAspAsnSer 770
 Db 2350 AAAGAGACTCTCAGAGGATTAATAAATAATTTACTATTGGCAGCCAGGACCAACAATACT 2409
 Qy 771 IleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer 785
 Db 2410 ATTATGGCAGAGCTGAGAAACTATTGGCTTTATTAAGGAGAGT 2454

RESULT 6
 US-09-468-656A-5
 ; Sequence 5, Application US/09468656A
 ; Patent No. 6582706
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-444
 ; CURRENT APPLICATION NUMBER: US/09/468,656A
 ; PRIOR FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 5
 ; LENGTH: 2531
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-09-468-656A-5

Alignment Scores:
 Pred. No.: 5,75e-266 Length: 2531
 Score: 2772.00 Matches: 541
 Percent Similarity: 75.78% Conservative: 91
 Best Local Similarity: 64.87% Mismatches: 142
 Query Match: 66.55% Indels: 60
 DB: 4 Gaps: 8

US-09-765-271-56 (1-796) x US-09-468-656A-5 (1-2531)
 Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19
 Db 61 TCTATGAATCTGGTCTGTCACCAAGTGGTCAAGTTAAGAAAGAGTCTAATCGAGTTTCT 120
 Qy 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGlnLeuThrProAspGluValSer 39
 Db 121 TATATAGATCGTATCAGGCTGGTCAAAAGCAGAAACCTTGACACAGATGAAGTCAGT 180
 Qy 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
 Db 181 AAGAGGAGGGGATCAACGCCGACAAATCGTCATCAAGATTACGGATCAAGGTTATGTG 240
 Qy 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIle 79
 Db 241 ACCTCTCATGGAGACCAATATCATTTACTATAATGGCAAGGTCCTTATGATGCATCATC 300
 Qy 80 SerGluGluLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
 Db 301 AGTGAAGAGCTCTCATGAAGATTCGAATATCATGTTGAAGGATTCAGACATTTGTCAT 360
 Qy 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119

Db 361 GAATCAAGGGTGGTTATGTATCAAGGTAGATGAGAAATACTATGTTTACCTTAAGGAT 420
 Qy 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
 Db 421 CGAGCTCATCGGATAATATTCGACAAAGAGAGATTAAACGTCAGAACGAGAACAC 480
 Qy 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
 Db 481 AGTCATTAATCACGGGGGTGGTTCT-----AACGATCAAGCAGTAGTTGCGAGCAGAGCC 534
 Qy 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
 Db 535 CAAGACGCTATACACGGATGATGGTTATATCTTCAATGCACTGATATCATTTGAGGAC 594
 Qy 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
 Db 595 ACGGGTGTATGCTTATATCGTTCCTCAGCGGACCATTAACCATTAACATTCCTAAGAATGAG 654
 Qy 200 LeuSerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
 Db 655 TTATCAGCTAGCAGATTAGCTGCTGCAGAAAGCCTATTGGAAATGGG----- 699
 Qy 220 AsnSerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239
 Db 700 -----AAGCAGGGATCTCGTCTCTTCAAGTCTCTAGTTATAATGCA 741
 Qy 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
 Db 742 AATCCAGCTCAACCAAGATTGTGAGAACCAACAATCTGACTGTCACTCCCAACTTATCAT 801
 Qy 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
 Db 802 CAA---AATCAAGGGGAAAAACATTTCAAGCCTTTTACGTGAATTTGATGCTAAACCCCTTA 858
 Qy 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
 Db 859 TCAGAACGCCATGTGGAATCTGATGGCTTTATTTTCGACCCAGCGCAATCACAAGTCGA 918
 Qy 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
 Db 919 ACCGCCAGAGGTAGTGTCTCTCATGGTAAACCATTAACCATTTATCTCTTATGAACAA 978
 Qy 320 MetSerGluLeuGluGluAlaArgIleAlaArgIleProLeuArgTyrArgSerAsnHis 339
 Db 979 ATGTCTGAATTGAAACCAAGATTTGCTGCTATTATTTCCCTTCTGTTATCTGTTCAAAACCAT 1038
 Qy 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
 Db 1039 TGGGTACCAGATTCAAGACCAGAACCAACCAAGTCCCAATCGACTCCGGAACCTAGTCCA 1098
 Qy 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373
 Db 1099 AGTCCGCAACTCGACCAAAATCTCAACCAAGTCTCAAGCAATCCAAATGAT-----GAG 1152
 Qy 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393
 Db 1153 AAATTGCTCAAGAGCTGTTTCGAAAGTAGGCGATGGTTATGCTTTGAGGAGAAATGGA 1212
 Qy 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413
 Db 1213 GTTTCGTATATCCAGCCAGGATCTTTTCAGCAGAAACACAGCAGCAGCATTTGATAGC 1272
 Qy 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433
 Db 1273 AAATGCCCAAGCAGGAAAGTTTATCTAAGCTAGGAGCTAAGAAACCTGACCTCCCA 1332
 Qy 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453
 Db 1333 TCTAGTGTGAGAAATTTTACAATAAGGCTTATGACTTACTAGCAAGAATTCACCAAGAT 1392
 Qy 454 LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473
 Db 1393 TTACTTGATAATAAAGTTCGACAAAGTTGATTTTGAGGCTTTGATACCTGTTGGAAACGA 1452

QY 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspLeuLeuAlaPheLeuAla 493
 Db 1453 CTCAGGATGTCCTCCAAAGTGAATAAGTCAAGTTAGTGATATATCTTGGCTTCTAGCT 1512
 QY 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluLysThrGluAsp 513
 Db 1513 CCGATTCGTATCCAGACGGTTAGGAAACCAATGCGCAATTTACCTACACATGATGAT 1572
 QY 514 GluValArgIleAlaGlnLeuAlaAspLysThrThrSerAspGlyTyrIlePheAsp 533
 Db 1573 GAGATTCAAGTAGCCAACTTGGCAGCAAGTACACACAGAGACGGTTATATCTTGTAT 1632
 QY 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553
 Db 1633 CCTCGTGATATAACCGAGTATGAGGGGATGCTATGTAATCTCCACATAGCCCTAGC 1692
 QY 554 HisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr 573
 Db 1693 CACTGGATTAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCAGCCAGGCTTATGCT 1752
 QY 574 LysGluLysGlyIleLeuProSerProAspAlaAspValLysAlaAsnProThrGly 593
 Db 1753 AAAGAGAAAGGTTTGCACCTCTTCGACAGACCATCAGGATTCAGGAATACTAGGCA 1812
 QY 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg 613
 Db 1813 AAAGAGCAGAGAGCTATCTACACCGCTGAAGCAGCTAAGAGAGTGCCTTATGCT 1872
 QY 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633
 Db 1873 ATGCCTTACAACTCTCAATATATCTAGAGTCAAAACGGTAGTTTATCATACCTCAT 1932
 QY 634 LysAspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAlaPro 653
 Db 1933 TATGACCATTAACATAACATAAATTTGAGTGGTTTGACGAGGCTTTATGAGCACT 1992
 QY 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673
 Db 1993 AAGGGGTATACTCTTGAGGATCTTTGGGACTGTCAAGTACTATGTGCAACATCAAC 2052
 QY 674 GluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLys--- 692
 Db 2053 GAACGTCGCAATTCAGATAATGTTTGGTAAACGCTAGCAGCACTGTTGTAATAAATAAG 2112
 QY 693 -----LysAspHisSerGlu----- 697
 Db 2113 GTAGACCAAGACAGTAACTGATGAAGATAAGAACATGATGAAGTAAAGTGAAGCAACT 2172
 QY 697 ----- 697
 Db 2173 CACCTCGAATCTGATGAAAGAGAAATCAGCGTGGTTTAAATCCTTCAGCAGATACTT 2232
 QY 698 ---AspProAsnLysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGlu 716
 Db 2233 TATAAACCAAGCACTGATACGGAAGAGACAGAGGAAGAGTCAAGATACCAAGATGAG 2292
 QY 717 ProGluValProGluValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluVal 736
 Db 2293 GCTGAAATTCCTCAAGTAGAATCTGTATTATTAACGCTTAAGATAGCAGATCGGAGGCC 2352
 QY 737 LeuLeuAlaLysValThrAspSerLeuLysAlaAsnAlaThrGluThrLeuAlaGly 756
 Db 2353 TTGCTAGAAAGAACTACAGATCTTAGTATTAGACAAATGCTATGGAGACATTGACTGGT 2412
 QY 757 LeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGlu 776
 Db 2413 CTAAAAAGTAGTCTTCTCTCGAACGAAAGATAATAACACTATTTCAGCAGAAAGTAGAT 2472
 QY 777 LysLeuLeuAlaLeuLysGlySerAsnProSerVal 790
 Db 2473 AGTCCTCTTGGCTTTGTTAAAGAAAGTCAACCGCTTCCTATA 2514

RESULT 7
 US-08-961-083-65
 ; Sequence 65, Application US/08961083
 ; Patent No. 6159469
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/961,083
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 65:
 ; LENGTH: 2290 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; US-08-961-083-65
 Alignment Scores:
 Pred. No.: 7,82e-254 Length: 2290
 Score: 2649.50 Matches: 519
 Percent Similarity: 76.51% Conservative: 77
 Best Local Similarity: 66.62% Mismatches: 130
 Query Match: 63.61% Indels: 53
 DB: 3 Gaps: 8
 US-09-765-271-56 (1-796) x US-08-961-083-65 (1-2290)
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19
 Db 5 TCTATGAACTTGGTCTGCACCAAGCTGGTCAGGTTAGAAAGAGTCTAATCGAGTTCT 64
 QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
 Db 65 TATATAGATGGTATCAGCTGGTCAAAAGGCGAGAAACTTCACACCAAGATCAAGTCACT 124
 QY 40 LysArgGluGlyLysAlaGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
 Db 125 AAGAGGAGGGGATCAACGCCGACAAATNGTNAATCAAGATTACCGATCAAGTTATG 184
 QY 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79
 Db 185 ACCTCTCATGGAGACCATATATCTATATATGCAAGGTTCCCTTATGATGCCATCATC 244
 QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
 Db 245 AGTGAAGAGCTCTCATGAAGATCCGAATTTATCATGTTGAAGATTGACACATTGTCAT 304
 QY 100 GluValLysGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119


```

Db 305 GAAATCAAGGTGTTATGTCATTAAGGTAAACCGTAAATCTATGNTACCTTAAGAT 364
Qy 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluLeuAsnArgGlnLysGlnGluHis 139
Db 365 GCAGCTCATGCGGATAATATCGGACAAAAGAGAGATTAACGTCAGAACGAGCAACGC 424
Qy 140 SerGlnHisArgGluGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
Db 425 AGTCATTAATCATCACTCA-----AGAGCAGATAATGCTGTGTGCGACGAGGCC 475
Qy 160 GlnGlyArgTyrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
Db 476 CAAGGAGTTATACAACGAGTATGGGTATATCTTCAATGATCATCTGATATCATTTAGGAC 535
Qy 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
Db 536 ACGGGTGATGCTTATATCGTTCTCCACGGCGACCATTAACCATTAACATTCCTAAGAAAG 595
Qy 200 LeuSerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
Db 596 TTATCACTAGCGAGTATGCTGCGAGAGCCATTGGAATGG----- 640
Qy 220 AsnSerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239
Db 641 -----AAGCAGGATCTGCTCTTCTCAAGTCTAGTTATAATGCA 682
Qy 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
Db 683 AATCCAGCTCAACCAAGATTGTAGAGAAACACCAATCTGACTGTCTCCCAACTTATCAT 742
Qy 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
Db 743 CAA---AATCAAGGGGAACATTTCAAGCCCTTTTACGTGAATGTATGCTAAACCCCTTA 799
Qy 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
Db 800 TCAGAACGCCATGTGGAAATCTGATGGCTTATTTTCGACCAGCGCAATCACAAGTCGA 859
Qy 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
Db 860 ACCGCCAGAGTGTAGTGTCTCCCTCATGGTAAACCATTAACCATTTATCTCCCTTATGAACAA 919
Qy 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
Db 920 ATGCTCAATTTGGAAGAACGAATTCGTGTATTAATCCCTTTCGTATCGTTCAAAACCAT 979
Qy 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
Db 980 TGGGTACCAGATTCACAGACCAACAACCAAGTCCACAAATCGACTCCGGAACTTAGTCCA 1039
Qy 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373
Db 1040 AGTCCGCAACCTTGCAACCAATCTCAACAGCTCCACAGCTCAACGAATCAATGTAT-----GAG 1093
Qy 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393
Db 1094 AAATGTTCAGAAAGCTGTTTCGAAAGTAGCGGATGTTATGCTTTTGAGGAGAAATGGA 1153
Qy 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413
Db 1154 GTTTCCTGTATATCCCGACCAAGATCTTTCAGCAGAAACAGCAGCAGGATTTGATGAGC 1213
Qy 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433
Db 1214 AAATGGCCCAAGCAGGAAAGTTTATCTCATAGCTAGGAGTAAAGAAACTGACCTCCCA 1273
Qy 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453
Db 1274 TCTAGTATCGAAGATTTTCAATTAAGCTTATGACTTACTAGCAAGAAATTCACCAAGAT 1333
Qy 454 LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473

```

```

Db 1334 TTACTTGATTAATAAGGTCGACAGTGTGATTTTGGAGCTTTGGATAACCTGTTGGAACGA 1393
Qy 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla 493
Db 1394 CTCAGGATGTGTCNCAAGTGTAAAGTCAAGTGTAGTGGANGATATCTTGCTCTTCTAGCT 1453
Qy 494 ProfileThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513
Db 1454 CCGATTCTCATCCAGAAACGTTTAGGAAACCAATGCGAAATACCTACACTGATGAT 1513
Qy 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533
Db 1514 GAGATTCAGTAGTCAAGATTGGCAGGCAAGTACACACAGAGACGGTATATCTTTGAT 1573
Qy 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553
Db 1574 CCTCGTATATAACAGTATGAGGGGATGCTTATGTAATCCACATATATGACCATAGC 1633
Qy 554 HistripIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr 573
Db 1634 CACTGGATTAAAAAGATAGTTTCTGTAAGCTGAGAGCGGCGAGCCAGGCTTATGCT 1693
Qy 574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593
Db 1694 AAAGAGAAAGTTTGACCCCTCTCTCGACAGACCATCAGGATTTCAGGAATACTGAGGCA 1753
Qy 594 AspSerAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg 613
Db 1754 AAAGGAGCAAGAGTATCTACACCGGTGAAAGCAGCTAAGAGGTGCCACTTGATCGT 1813
Qy 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633
Db 1814 ATGCTTACATCTTCAATATCTGTAGAAGTCAAAACCGTGTATTAATCATACCTCAT 1873
Qy 634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653
Db 1874 TATGACCATTTACCATTAACATCAATTTGAGTGTGTTCAGGAGCCCTTTATGAGGCACCT 1933
Qy 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673
Db 1934 AAGGGGTATATCTCTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAAC 1993
Qy 674 GluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLys 693
Db 1994 GAACGTCGCGCATTCAGATAATGTTTGGTAAACGCTAGCGACCATGTGTCAAAGAAACAAA 2053
Qy 694 AspHisSerGluAspProAsnLysAsnPheLysAlaAspGluGlu----- 708
Db 2054 AATGCTCAAGCTGATACCAATCAAAACGAAACCAACGAGCGAGGAGAAACCTCAGACAGAA 2113
Qy 709 ---ProValGluGluThr----- 713
Db 2114 AAACCTGAGGAAGAAACCCCTCGAGAGAGAAACCGCAAGCGAGAACCCAGAGTCTCCA 2173
Qy 714 -----ProAlaGluProGluValProGlnValGluThr 724
Db 2174 AAACCAACAGAGAACACAGAGAAATCCAGAGGAATCAGAGAACCTCAGTCCGAGACT 2233
Qy 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAsp 743
Db 2234 GAAAAGGTTGAAGAAAACTGAGAGAGGCTGAAGATTACTTGGAAAATATCCAGAT 2290

```

RESULT 8

US-09-536-784-65

; Sequence 65, Application US/09536784

; Patent No. 6573082

; GENERAL INFORMATION:

; APPLICANT: Choi et. al.

; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

; NUMBER OF SEQUENCES: 452

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 65:
SEQUENCE CHARACTERISTICS:
LENGTH: 2290 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 65:
US-09-536-784-65

Alignment Scores:
Pred. No.: 7,82e-254 Length: 2290
Score: 2649.50 Matches: 519
Percent Similarity: 76.51% Conservative: 77
Best Local Similarity: 66.62% Mismatches: 130
Query Match: 63.61% Indels: 53
DB: 4 Gaps: 8

US-09-765-271-56 (1-796) x US-09-536-784-65 (1-2290)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19
DB 5 TCCTATGAACCTGGTTCCTCACCACAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTCT 64
QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
DB 65 TATATAGATGGTGATCAGCTGGTCAAAAGGCGAGAAACTTGACACCCAGATGAAGTCAGT 124
QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
DB 125 AAGAGGAGGGGATCAACCCCAACAAATNGNATCAAGATTACGATACAGGTTATGTG 184
QY 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIlelle 79
DB 185 ACCTCTCATGGAGACCATATCATCTACTATATGCAAGGTTCTTATGATCCCATCATC 244
QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
DB 245 AGTGAAGAGCTCCTCATGAAGATCCGAATTATCATGTTGAAGATTTCAGACATTGTCAT 304
QY 100 GluValLysGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
DB 305 GAAATCAAGGGTGGTATGTCATTAAGGTAACGGTAATATACTATGTTACCTTAAGGAT 364
QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
DB 365 GCAGCTCATCGGATAATATTCGACAAAGAGAGATTAACACGTCAGAAAGCAGGAAACGC 424
QY 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
DB 425 AGTCATATCATAACTCA-----AGACGAGATATGCTGTGCTCGACCCAGAGCC 475

QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
DB 476 CAAGACGTTTATACAAACGATGGTATATCTTCAATGATCTGATATCATTTAGGAC 535
QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
DB 536 ACGGGTGATGTTATATCTGTTCTTCCACGGGACCATTACCATTAATCTTAAAGATGAG 595
QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
DB 596 TTATCAGCTAGCGAGTTAGCTGCTGCAGAAAGCTTATGGAATGGG----- 640
QY 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239
DB 641 -----AAGCAGGATCTCGTCTCTTCTCAAGTCTAGTTTATATGCA 682
QY 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
DB 683 AATCCAGCTCAACCAAGATTGTGCAGAAACCAACATCTGACTGTCTCACTCAACTTATCAT 742
QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
DB 743 CAA---AATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTGTATGCTTAAACCCCTTA 799
QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
DB 800 TCAGAACCCATGTGAATCTGATGCGCTTATTTGACCCAGCCCAATCAAGTCA 859
QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
DB 860 ACCGCCAGAGGTGTAGCTGCTCCTCATGTAAACCATTACCATTTATCCCTTATGAACAA 919
QY 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
DB 920 ATGCTGAATTGGAATAACGAATTGCTCGTATTATTCCTCTTCGTTATGCTTCAACCAT 979
QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
DB 980 TGGGTACAGATTCAAGACCAGAACCAACAGTCCACCAATCGACTCCGGAACCTAGTCCA 1039
QY 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373
DB 1040 AGTCGCAACCTCGACCAAAATCTCAACAGCTCCCAAGCAATCCAAATTCAT-----GAG 1093
QY 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393
DB 1094 AAATTTGGTCAAGAAAGCTGTTGCAAAAGTAGCGATGTTATGTTCTTTGAGGAGATGGA 1153
QY 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413
DB 1154 GTTCTCTGTTATATCCAGCCAGGATCTTTCAGCAGAAACAGCAGCAGGATTCATAGC 1213
QY 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysGlyLysGluAsnValAla 433
DB 1214 AAATGCGCAAGCAGGAAAGTTTATCTCATAGCTAGGAGCTAAGAAACCTACACCTCCA 1273
QY 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453
DB 1274 TCTAGTATCAGAAATTTTACAAATAGGCTTATGATTTACTAGCAGAAATTCACCAAGAT 1333
QY 454 LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473
DB 1334 TTACTTGATATAAAGGTCGACCAAGTTGATTTTGAGGCTTTTGATAAACCCTGTGGAACA 1393
QY 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla 493
DB 1394 CTCAGAGGATGTCNCAGGTATAAAGTCAAGTTAGTGGAGATATTTCTTGCCTCTTAGCT 1453
QY 494 ProIleThrHisProGluLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513
DB 1454 CGATTCTCATCCAGAACGTTTAGAAAAACCAATGCGCAATTTACCTACACTGATCAT 1513

Qy	514	GluValArgIleAlaGlnLeuAlaAspLysTyrThrSerAspGlyTyrIlePheAsp	533
Db	1514	GAGATTCAAGTAGGCAAGTTGGCAGGCAAGTACACAAACAGACGGTTATATCTTTGAT	1573
Qy	534	GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer	553
Db	1574	CCTCGTGATATACCACTGATGAGGGGGATGCCCTATGTAACTCCACATATGACCCCATAGC	1633
Qy	554	HieTriPileGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr	573
Db	1634	CACCTGGATTAAAAAGATAGTTTGCTCGAGCTGAGAGCGGCAGCCAGGCTTATGCT	1693
Qy	574	LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAenProThrGly	593
Db	1694	AAAGAGAAAGGTTTGACCCCTCTCTTCACAGACCATCAGGATTACAGAAATACCTGAGGCA	1753
Qy	594	AspSerAlaAlaAlaIleTyrAsnArgValLysGlyLulysArgIleProLeuValArg	613
Db	1754	AAAGGAGCAGAGCTATCTCAACCCGGTGAAGACAGCTAAGAAAGGTGCCACTTGATCGT	1813
Qy	614	LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis	633
Db	1814	ATGCCITACAATCTTCAATATATCTGTAGAGTCAAAAACGGTAGTTTAATCATACCTCAT	1873
Qy	634	LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro	653
Db	1874	TATGACCATTACCATAAACATCAAAATTTGAGTGGTGTTCACCAAGGCCTTTATGAGGCACCT	1933
Qy	654	AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp	673
Db	1934	AAGGGGTATACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTTCGAACATCCAAAC	1993
Qy	674	GluArgProHisSerAsnAspGlyTyrPGLysAlaSerGluHisValLeuGlyLysLys	693
Db	1994	GAAGCTCCGATTTCAGATAATGTTTGGTAAAGCTAGCGACCATGTTCAAAGAAACAA	2053
Qy	694	AspHisSerGluAspProAsnLysAsnPheLysAlaAspGluGlu	708
Db	2054	AATGGTCAAGCTGATACCAATCAACCGGAAAAACCAAGCGAGGAGAAACCTCAGACAGAA	2113
Qy	709	---ProValGluGluThr	713
Db	2114	AAACCTGAGGAAGAAACCCCTCGAGAAGAGAAACCGCAAGCGAGAGAAACCGAGAGTCTCCA	2173
Qy	714	-----ProAlaGluProGluValProGlnValGluThr	724
Db	2174	AAACCAACAGAGAACCCAGNAGAAATCACCGAGGAATCAGAGAACTCAGGTCGAGACT	2233
Qy	725	GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlalysValThrAsp	743
Db	2234	GAAGAAGTTGAAGAAAACTGAGAGAGGCTGAAGATTCTTGGAAAAATCCAGGAT	2290

RESULT 9

US-08-961-527-243

; Sequence 243, Application US/08961527

; Patent No. 6420135

; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 391

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

Db 1531 TTATCAGCTAGCGAGTGTAGCTGCTGACAGACGCTATTGGAATGGG----- 1575
QY 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239
1576 -----AAGCAGGAGTCTCGTCTCTTCAAGTCTTAGTTATAATGCA 1617
QY 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
1618 AATCAGCTCAACAGAGATTGTCAGAGAACACAACTGACGTGCACCTCAACCTATCAT 1677
QY 260 GlnAlaSerGlnSerAsnAspSerLeuLeuLeuGlnLeuLeuLeuLeuLeuLeuLeuLeu 279
1678 CAA---AATCAAGGGGAAACATTTCAAGCCCTTTACGTGAATTTGATGCTAAACCCCTA 1734
QY 280 SerGlnArgHisValGlnSerAspGlyLeuValPheAspProAlaGlnLeuLeuLeuLeuLeu 299
1735 TCAGAACCCCATGTTGGAATCTGATGCGCTTATTTTCGACCCGACCAATCACAAGTGA 1794
QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
1795 ACCGCCAGAGGTGAGTGTCCCTCATGGTAACCATTACCCTTTATCCCTTATGAACA 1854
QY 320 MetSerGluLeuGluGluArgGlnLeuAlaGlnLeuLeuLeuLeuLeuLeuLeuLeuLeu 339
1855 ATGCTGAATTTGGAAGAACGAATTCGTGATTTATTCCTTCGTTATCGTTCAACACCAT 1914
QY 340 TrpValProAspSerArgProGlnProGlnProSerProGlnProThrProGluProSerPro 359
1915 TGGGTACCAGATTCAAGACCAAGAACCAAGTCCACACCGACCTCCAGAACCTAGTCCA 1974
QY 360 GlyProGlnProAlaProAlaProAlaProAlaProAlaProAlaProAlaProAlaPro 379
1975 ACTCGCAACCACTCCAGCAATCAATTTGAT-----GACAAATTTGTCACCAAGCAAGCT 2028
QY 380 ValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPhe 399
2029 GTTCGAAAGTAGGCGCTTATGTTTCTTTGAGGAGAAATGAGTTCCTGTTATATCCCA 2088
QY 400 AlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGlu 419
2089 GCCAAGGATCTTCAGCAGAAACAGCAGCAGCATTTGATGACAACTGGCCCAAGCAGGA 2148
QY 420 SerValSerHisThrThrAlaLysLysGluAsnValAlaProAlaProAlaProAlaPro 439
2149 AGTTTATCTCAAGCTAGGAACTAAGAACTGACCTCCCATCTAGTGCAGAAATTT 2208
QY 440 TyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGly 459
2209 TACAATAAGGCTTATGACTTACTAGCAAGAATTCACCAAGATTTACTTGAATAAAGGT 2268
QY 460 ArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThr 479
2269 GCACAAGTTGATTTTGGAGCTTTGGATACCTTTGGAAACGACTCAAGATGCTCAAGT 2328
QY 480 AsnLysGluLysLeuValAspLeuLeu 489
2329 GATAAAGTCAAGTTAGTGAAGATATCTTT 2358

RESULT 10

US-09-765-271-56 (1-796) x US-09-765-271-56 (1-6867)
; Sequence 192, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 192:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6867 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-765-271-56 (1-796) x US-09-765-271-56 (1-6867)

Alignment Scores:
Pred. No.: 6,768-111 Length: 6867
Score: 1222.50 Matches: 295
Percent Similarity: 54.22% Conservative: 97
Best Local Similarity: 40.80% Mismatches: 181
Query Match: 29.35% Indels: 151
DB: 4 Gaps: 18

US-09-765-271-56 (1-796) x US-09-765-271-56 (1-6867)
QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db 6722 GCCTATGCATTAACCCAGCATCGTTTCGACGAGGAAAAAAGGACAAATAATCGTCTCTTAT 6663
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 6662 GTGGATGCGCAGCATCAAGTCAGAAAAGTGAACCTTGACACCCAGACGAGTTAGCCAG 6603
QY 41 ArgGluGlyIleAsnAlaGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db 6602 AAGAAGGAATTCAGGCTGAGCAAAATTTGATCAAAATTCAGATCAGGCTATGTAAAG 6543
QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleLeuSer 80
Db 6542 TCACACGGTGACCATCTATCTATTAATGGGAAAGTTCTTATGATGCCCTCTTTAGT 6483
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db 6482 GAAGAAGTCTTGATGAAGGATCCAACTATCAACTTAAAGACGCGTATGTCATGAAG 6423
QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysLysAspAla 120
Db 6422 GTCAAGGGTGGTTATATCATCAAGTCGATGGAATAATTTATGCTACTCTGAAGATGCA 6363
QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
Db 6362 GCTCATGCTGATAATGTTTCGAATAAGATGAATCAATCGTCAAAAACAAGAACATGTC 6303
QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
Db 6302 AAGAATAATGAG-----AAGGTTAACTTAATGTTGCTGACCAAGCTCAG 6255
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db 6254 GGACCATATACCAAAATGATGTTTATGCTTTTAAATCCAGCTGATATTATCAAGATACG 6195

QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
 DB 6194 GGTAAATGCTTATATCGTTCTCATGAGCTCACTATCACTATCCCAAGCAAGCAATTA 6135
 QY 201 SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
 DB 6134 TCTGCTAGTGAATAGCAGCAGCTAAAGCACAATCTGGCTGGAAAAAATATGCAACCGAGT 6075
 QY 221 SerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
 DB 6074 CAGTTAAGCTATTCTTCAACAGCTAGTGCAT 6042
 QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
 DB 6041 -----AACAGCAATCTGTAGCAAAAGGATCAACTAGCAAGCCA 6003
 QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLysSer 280
 DB 6002 GCAATAAATCTGAATAATCTCCAGAGCTCTTTGAAGGAACCTATGATTCACCTAGCGCC 5943
 QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
 DB 5942 CAACGTTACAGTGAATCAGATGGCTGTCTTTGACCTGCTAGATATCAGTCGTACA 5883
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
 DB 5882 CCAATGGAGTTGGCAATTCGCATGGCGACCACTTACCACCTTTATTCCTTACAGCAAGCTT 5823
 QY 321 SerGluLeuGluAlaGlyIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340
 DB 5822 TCTGCCCTTAGAAGAAAGATTGCGAGATGGCGCT 5787
 QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
 DB 5787 ----- 5787
 QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
 DB 5786 -----ATCAGTGGAACTGGTTCTCAGATTTCTACAAATGCA 5751
 QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
 DB 5750 AAA----- 5748
 QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
 DB 5747 -----CCTAATGAAGTAGTGTCTAGCTAGGCACTTTTCAACCAATCCTCTTCT 5697
 QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
 DB 5696 -----TTAACGACAAGTAAGGAGCTCTTTCAGCATCTGATGTTATATTTT 5649
 QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
 DB 5648 AATCCAAA-----GATATCGTTGAAGAACCGCTACAGCTTATATGTAAGACATGCTG 5595
 QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
 DB 5594 ---GATCAATTCCTACATCCAAA-----TCAATCAAAATGGG 5556
 QY 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
 DB 5555 CAACCGACTCTCCAAACAATAGTCTAGCAACACCTTCTCCA---TCTCTTCCAAATCAAT 5499
 QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
 DB 5498 CAGGAACCTTCAATGAGAAACATGAA----- 5472
 QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
 DB 5471 -----GNAAGATGATACGGAATTTGATGCTAATCGTATATTCGCTGAA 5430
 QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560

DB 5429 GATGAATCAGGTTTTTGTCTAGTCCAGGACCAACATCATTTCTTCAAGAGGAC 5370
 QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
 DB 5369 TTGACAGAAGACAAATTA--GSTGCGCAAAACATTTAGAGAA----- 5326
 QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600
 DB 5325 -----GTTAAAACTAGTATATGATGATTTCTTTGTCTATCTCAT 5284
 QY 601 AsnArg-----ValLysGlyGluLys 607
 DB 5283 GAAACAGGATTATCCAGTAATGCCAAGAAATGAAGATTAGATAAAAAAATCGAAGAA 5224
 QY 608 ArgIleProLeuValArgLeuProTyrMetVal-----GluHisThrValGluValLys 625
 DB 5223 AAAATTTGCTGCATTTATGAAACAATATGTTGCTCAACCTGAAAGTATTGTCGTGAATAA 5164
 QY 626 AsnGlyAsnLeuIleIle---ProHisLysAspHisTyrHisAsnIleLysPheAlaTyr 644
 DB 5163 GAAAAAATGCGATTATTATTCGCGATCGAGATCACCATCATCGAGATCCG----- 5113
 QY 645 PheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAlaThr 664
 DB 5112 ATTGATGAACAT-----AAACCGTTGGAATTGTT-----CATTTCTCAC 5074
 QY 665 IleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAspGly----- 681
 DB 5073 AGTAACTATGAACCTGTTTAAACCGAAGAGGAGTTGCTTAAAGAGGGAATAAAGTT 5014
 QY 682 ---TrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
 DB 5013 TATACTGGAGAAGATTACGAATGTTGTTAATTGTTTAAATAATAGTACGTTTAAATAAT 4954
 QY 701 LysAsnPhe 703
 DB 4953 CAAACCTTT 4945
 RESULT 11
 US-09-468-656A-7
 ; Sequence 7, Application US/09468656A
 ; Patent No. 6582706
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-444
 ; CURRENT APPLICATION NUMBER: US/09/468,656A
 ; CURRENT FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 7
 ; LENGTH: 1455
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-09-468-656A-7
 Alignment Scores:
 Pred. No.: 4.28e-110 Length: 1455
 Score: 1203.50 Matches: 259
 Percent Similarity: 59.47% Conservative: 77
 Best Local Similarity: 45.84% Mismatches: 120
 Query Match: 28.90% Indels: 109
 DB: 4 Gaps: 10
 US-09-765-271-56 (1-796) x US-09-468-656A-7 (1-1455)
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20

```
Db 64 GCCATGACATTAACAGCATCGTTCGAGAAATAAGACATAATCGTCTCTTAT 123
Qy 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 124 GTGGATGCGCAGCCAGTCAAGTCAGAAAAGTGAACACTTGACACACAGAGCGGTAGCCAG 183
Qy 41 ArgGluGlyIleAsnAlaGluGlnIleValLysIleThrAspGlnGlyTyrValThr 60
Db 184 AAAGAAGGAATTCAGGCTGAGCAAAATGTAATCAAAATTTACAGATCAGGGCTATGTAACG 243
Qy 61 SerHisGlyAspHisTyrThrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db 244 TCACAGGTGACCATCATCTATCTATTAATGGAAAGTTCCTTATGATGCCCTCTTTAGT 303
Qy 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db 304 GAAGAACTCTTGATGAAGGATCCAACTATCAACTTAAAGACGCTGATATGTCATGA 363
Qy 101 ValLysGlyGlyTyrValLysValLysValAspGlyLysTyrThrValTyrLysAspAla 120
Db 364 GTCAAGGGTGGTATATCATCAAGGTCGATGGAATAATATTGTCTCTGAAAGATGCA 423
Qy 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnHisSer 140
Db 424 GCTCATGCTGATATGTCGAACCTAAAGATGAATCAATCGTCAAAACAAGACATGTC 483
Qy 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
Db 484 AAAGATAATGAG-----AAGGTTAACTCTAATGTTCGTGTAGCAAGGCTCTCAG 531
Qy 161 GlyArgTyrThrThrAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db 532 GGACATATACCAACAATGATGTTATGCTTTAATCCAGCTGATATTAATCGAAGATACG 591
Qy 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db 592 GGTAAATGCTTATATCGTTCCTCATGGAGTCACTATCACTATCCCAAAAAGCGATTTA 651
Qy 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db 652 TCTGCTAGTGAATTAGCAGCAGCTAAAGCACATCTGGCTGGAAAAAATATGCAACCGAGT 711
Qy 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
Db 712 CAGTTAAGCTATTCTTCAACAGCTAGTGCAAT----- 744
Qy 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db 745 -----AACACGCAATCTGTAGCAAAAGGATCAACTAGCAAGCCA 783
Qy 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
Db 784 GCAAAATAAATCTGAAAATCTCCAGAGTCTTTGAAGGAAGTCTATGATTCACCTAGCGCC 843
Qy 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db 844 CAACGTTACGTGAATCAGATGCGCTGGTCTTTGACCTGCTAAGATTATCAGTCGTACA 903
Qy 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db 904 CCAATGGAGTTCGATTCGCGATGCGCACCATACCTTTATTCCTTACAGCAAGCTT 963
Qy 321 SerGluLeuGluGluArgIleAlaArgIlePheProLeuArgTyrArgSerAsnHisTrp 340
Db 964 TCTGCTCTAGAGAAAAGATTGCGAAGTGGTGCT----- 999
Qy 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db 999 ----- 999
Qy 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
```

```
Db 1000 -----ATCAGTGGAACTGGTTCTTACAGTTCTTACAAATGCA 1035
Qy 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
Db 1036 AAA----- 1038
Qy 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
Db 1039 -----CCTAATGAAGTAGTGTCTAGTCTAGGCAGTCTTTCAAGCAATCTTCTTCT 1089
Qy 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
Db 1090 -----TTAAACGACAAAGTAAGAGCTCTTCCAGCATCTGATGTTATATTTT 1137
Qy 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db 1138 AATCCAAAA---GATATCGTTCAAGAAACGGCTACAGCTTATATTGTAAGACATGCT-- 1191
Qy 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
Db 1192 ---GATCATTTCCATTACATTCCAAAA-----TCAAATCAAATGGG 1230
Qy 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db 1231 CAACCGACTCTCCAAACAATAGTCTAGCAACACCTTCTCCA---TCTCTCCAATCAAT 1287
Qy 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
Db 1288 CCAGAACTTCATCAGAAACATGAA----- 1314
Qy 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
Db 1315 -----GAAGATGGATACGATTGATGCTAATCGTATTCGCTGAA 1356
Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560
Db 1357 GATGAATCAGGTTTGTGTATGATGTCAGGAGACCAATCAATATTATTCTTCAGAGGAC 1416
Qy 561 LeuSerAspLysGlu 565
Db 1417 TTGACAGAGAGCAA 1431
```

RESULT 12
US-08-961-083-181
; Sequence 181, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2

; ;
; ;
; ;
; ;
; ;
; ;

CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland

```

;
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
;
; INFORMATION FOR SEQ ID NO: 181:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1342 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 181:
US-09-536-784-181

```

```

Alignment Scores:
Pred. No.: 1,47e-108 Length: 1342
Score: 1187.50 Matches: 257
Percent Similarity: 59.28% Conservatives: 72
Best Local Similarity: 46.31% Mismatches: 117
Query Match: 28.51% Indels: 109
DB: 4 Gaps: 10

```

US-09-765-271-56 (1-796) x US-09-536-784-181 (1-1342)

```

QY 4 LeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyrIleAspGly 23
DB 2 CTTAAACCCAGCATCGTTCCGAGGAAATAAGGCAATAATCGTGTCTCTTTATGTGGATGC 61
QY 24 LysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLysArgGluGly 43
DB 62 AGCCAGTCAGTCAGAAAGTGAACACTTGACACAGACCCAGGTAGCCGAGAAAGGA 121
QY 44 IleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThrSerHisGly 63
DB 122 ATTCAGGCTGAGCAATTTGTAATCAAAATTTACAGATCAGGGCTATGTAACGTCACACGGT 181
QY 64 AspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSerGluGluLeu 83
DB 182 GACCACTATCATTTACTATAATGGGAAAGTTCCTTATGATGCCCTCTTTATGGAAGAACTC 241
QY 84 LeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGluValLysGly 103
DB 242 TTGATGAAGATCCAAACTATCACTTAAAGACGCTGATATTGTCAATGAAGTCAAGGT 301
QY 104 GlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysAspAlaIleHisAla 123
DB 302 GGTATATCATCAAGTCGATGAAATATTATGCTACTCTGAAAGATCGAGTCTATGCT 361
QY 124 AspAsnValArgThrLysGluLysIleAsnArgGlnLysGlnGluHisSerGlnHisArg 143
DB 362 GATAATGTTCCGAACCTAAGATGAATCAATCGTCAAAACAAAGAAATGTCAAAGATAAT 421
QY 144 GluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGlnGlyArgTyr 163
DB 422 GAG-----AAGGTTAACTCTAATGTTGCTGTAGCAAGGCTCAGGACCATAT 469
QY 164 ThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThrGlyAspAla 183

```

```

DB 470 ACGCAAAATGATGGTTATGTCTTTAATCCAGCTGATATTATCGAAGATACGGTAATGCT 529
QY 184 TyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeuSerAlaSer 203
DB 530 TATATCGTTCTCTATGGAGGTCACATCATCATCTCCAAAGGATTTATCTCTAGT 589
QY 204 GluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSerArgThr 223
DB 590 GAATTAGCAGGAGCTAAAGCACTCTGGCTGGAAAAAATATGCAACCGAGTCAGTTAAGC 649
QY 224 TyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSerValSerAsn 243
DB 650 TATCTTCAACAGCTAGTGACAAT----- 673
QY 244 ProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAlaSerGln 263
DB 674 -----AACACGCAATCTGTAGCAAAAGATCACTAGCAGCCCAATATAA 721
QY 264 SerAsnAspIleAspSerLeuLysGlnLeuTyrLysLeuProLeuSerGlnArgHis 283
DB 722 TCTGAAATCTCCAGAGTCTTTTGAAGGAACCTCTATGATTCACTAGCGCCCAAGCTTAC 781
QY 284 ValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThrAlaArgGly 303
DB 782 AGTGAATCAGATGGCTGCTCTTTGACCTGCTAAGATTATCAGTCGTACACCAATGGA 841
QY 304 ValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMetSerGluLeu 323
DB 842 GTTGGATTCCGCATGGCGACCAITTACCACCTTTATCTTACAGCAAGCTTCTGCCCTTA 901
QY 324 GluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrpValProAsp 343
DB 902 GAAGAAAGATTGCCAGATGGTGCCT----- 928
QY 344 SerArgProGluGlnProSerProGlnProThrProGluProSerProGlyProGlnPro 363
DB 928 ----- 928
QY 364 AlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuValArgLysVal 383
DB 929 -----ATCAGTGGAACTGGTCTTACAGTTCTACAAATCAAA----- 967
QY 384 GlyGluGlyTyrValPheGluGlyLysGlyIleSerArgTyrValPheAlaLysAspLeu 403
DB 967 ----- 967
QY 404 ProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSerValSerHis 423
DB 968 CCTAATGAAGTAGTGTCTAGTCTAGGCACTCTTCAAGCAATCCTTCTTCT----- 1018
QY 424 ThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyrAspLysAla 443
DB 1019 -----TTAAGCAAGTAAGAGCTCTCTCAGCATCTGATGGTTATATTTTATCCAAAA 1075
QY 444 TyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArgAsnSerAsp 463
DB 1076 ---GATATCGTTGAAGAAACCGCTACAGCTTATTTTGAAGACATGTT-----GATCAT 1126
QY 464 PheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsnLysGluLys 483
DB 1127 TTCCATTATCATTTCCAAAA-----TCAATCAATTCGGCAACCGACT 1168
QY 484 LeuValAspLeuLeuAlaPheLeuAlaPheIleThrHisProGluArgLeuGlyLys 503
DB 1169 CTTCCAAACAATAGTCTAGCAACACCTTCTCCA---TCTCTTCCAATCAATCCAGGAAC 1225
QY 504 ProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeuAlaAspLys 523
DB 1226 TCACATGAGAAACATGAA----- 1243
QY 524 TyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAspGluGlyAsp 543

```



```

Db 1244 -----GAAGATGGATACGGATTTCATCTAATCGTATTCGCTGAAGATCAATCA 1294
Qy 544 AlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLys 558
Db 1295 GGTTCATGAGTCACGGAGACCAATCATTTATTCCTCAAG 1339

RESULT 14
US-08-961-527-355
; Sequence 355, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 355:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-355

Alignment Scores:
Pred. No.: 6,57e-89 Length: 973
Score: 988.00 Matches: 203
Percent Similarity: 79.79% Conservative: 26
Best Local Similarity: 70.73% Mismatches: 42
Query Match: 23.72% Indels: 17
DB: 4 Gaps: 4

US-09-765-271-56 (1-796) x US-08-961-527-355 (1-973)

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnArgValSer 19
Db 158 TCCTATGAGCTTGGACGCTTACCAAGCTGGTCAGGATAAGAAAGAGTCTAATCGAGTTGCT 217
Qy 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
Db 218 TATATAGATGCTGATCAGGCTGGTCAAAAGGCGAGAAACCTTGACACAGATGAAGTCAGT 277
Qy 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
Db 278 AAGAGGAGGGGATCAACGGCGACAAATTTGTTATCAAGATTACGGATCAAGTTATGTCG 337
Qy 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIle 79
Db 338 ACCTCTCATGGAGACCATTTATCTACTATAATGGCAAGGTTCTTATGATGCAATCATC 397

```

```

Qy 80 SerGluGluLeuLeuMetLysAspProAsnTyrIleLysLeuLysAspGluAspIleValAsn 99
Db 398 AGTGAAGAGCTCCTCATGAAGAGATCCGAATATATCATGTTGAAGGATTTCAGACATTTGTCAAT 457
Qy 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
Db 458 GAAATCAAGGGTGGTTATGTCATTAAGGTAAACCGTAAATACTATGTTTACCTTAAGGAT 517
Qy 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
Db 518 GCAGCTCATGCGGATAATATTCGGACAAAGAGAGATTAAACGCTCAGAAGCAGGAAACGC 577
Qy 140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
Db 578 AGTCATAATCATAACTCA-----AGACCAAGATAATGCTGTGTGTCAGCCAGAGGCC 628
Qy 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
Db 629 CAAGGACGTTATACACGAGATGATGGGTATATCTTCAATGCATCTGATATCATTTAGGAC 688
Qy 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
Db 689 ACGGTGATGCTTATATCGTTCCTCAGCGGACCATTAACCATTAACATTCTTAAGAATGAG 748
Qy 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
Db 749 TTATCAGCTAGCGAGTTAGCTGCTGCAGAACGCTTATTGGAATGGG----- 793
Qy 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239
Db 794 -----AAGCAGGATCTCGCTCTTCTCAAGTCTCTGATTATAATGCA 835
Qy 240 SerValSerAsnProGlyThrThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
Db 836 AATCCAGCTCAACCAAGATTGTCAGAGAACCAACATCTGACTGCTCACTCCAACTTATCAT 895
Qy 260 GlnAlaSerGlnSerAsnAspIleAspSerIleLeuLysGlnLeuTyrLysLeuProLeu 279
Db 896 CAA---AATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTTGTATGCTAA-CCCTTA 951
Qy 280 SerGlnArgHisValGluSer 286
Db 952 TCAGACGCCCATGTGGGATCT 972

RESULT 15
US-08-961-527-258
; Sequence 258, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

```

NAME: Brookes, A. Anders
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PB340P1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 258:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1684 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 US-08-961-527-258

Alignment Scores:
 Pred. No.: 7,448-67 Length: 1684
 Score: 770.50 Matches: 157
 Percent Similarity: 67.90% Conservative: 27
 Best Local Similarity: 57.93% Mismatches: 58
 Query Match: 18.50% Indels: 30
 DB: 4 Gaps: 2

US-09-765-271-56 (1-796) x US-08-961-527-258 (1-1684)

QY 544 AlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSerLeuSerAsp 563
 DB 3 GCCTATGTAACCTCCACATATGACCATAGCCATGACCTGATTAAGAAAGATGTTGCTGAA 62
 QY 564 LysGluLysValAlaGlnAlaTyrThrLysGluLysGlyIleLeuProSerPro 583
 DB 63 GCTGAGAGCGCGACC-CAGGCTTATGCTAAAGAGAAAGTTTGACCCCTCCTTCGACA 121
 QY 584 AspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyrAsnArgVal 603
 DB 122 GACCATAGGANTCGAGAACTACTGAGGCAAGGAGCAGAGCTATCTACACCGCGTG 181
 QY 604 LysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHisThrValGlu 623
 DB 182 AAAGCAGCTAAGAGGTCGACCTTGATCGTATGCCTTACAACTTCAATATATCTAGAA 241
 QY 624 ValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIleLysPheAla 643
 DB 242 GTCAAAAACGGTAGTTAATCATACCTCATATGACCATACCAATACCAATCAAAATTTGAG 301
 QY 644 TrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAla 663
 DB 302 TGGTTTGACGAGGCTTTATGAGGACCTAAGGGGTATCTCTGAGGATCTTTGGCG 361
 QY 664 ThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAspGlyTrpGly 683
 DB 362 ACTGTCAGTACTATGTGAAACATCCAAACGAAACGTCGCGCATTCAGATAATGGTTTGGT 421
 QY 684 AsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsnLysAsnPhe 703
 DB 422 AACGCTAGCGACCATGTTCAAGAGAAACAAATAATGGTCAAGCTCATACCAATCAAAACGAA 481
 QY 704 LysAlaAspGluGlu-----ProValGluGluThr----- 713
 DB 482 AAACCAAGCGAGGAGAAACCTCAGACAGAAACCTTGAGAGAAACCCCTCAGAGAGAG 541
 QY 714 -----Pro 714
 DB 542 AAACCGCAAGCGAGAAACAGAGTCTCCAAAACCAACAGAGAAACAGAGAAATCAACCA 601
 QY 715 AlaGluProGluValProGlnValGluThrGluLysValGluAlaGlnLeuLysGluAla 734
 DB 602 GAGGAATCAGAGAAACCTCAGTCTGAGACTGAAAGAGTTGAAGAAATACTGAGAGAGCT 661
 QY 735 GluValLeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeu 754
 DB 662 GAAGATTACTTGGAAAAATCCAGATCCAAATATCAAGTCCAAATCCAAAGAGACTCTC 721
 QY 755 AlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGlu 774

DB 722 ACAGGATTAAAAATAATTACTATTTGGCACCAGGACAACTACTATTATGCGAGAA 781
 QY 775 AlaGluLysLeuLeuAlaLeuLeuLysGlySer 785
 DB 782 GCTGAAAAAATCTATTGGCTTTTATTAAAGGAGAGT 814

RESULT 16

US-08-743-637B-34/C
 ; Sequence 34, Application US/08743637B
 ; Patent No. 5994066

GENERAL INFORMATION:

APPLICANT: BERGERON, Michel G.
 APPLICANT: PICARD, Francois J.
 APPLICANT: QUELLETTE, Marc
 APPLICANT: ROY, Paul H.

TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
 TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
 TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
 TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
 NUMBER OF SEQUENCES: 273
 CORRESPONDENCE ADDRESS:

ADDRESSEE: QUARLES & BRADY
 STREET: 411 EAST WISCONSIN AVENUE
 CITY: MILWAUKEE
 STATE: WISCONSIN
 COUNTRY: USA
 ZIP: 53202-4497

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/743,637B
 FILING DATE: 04-NOV-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/526,840

FILING DATE: 11-SEP-1995

ATTORNEY/AGENT INFORMATION:

NAME: BAKER, Jean C.

REGISTRATION NUMBER: 35,433

REFERENCE/DOCKET NUMBER: 850586,90012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (414) 277-5000

TELEFAX: (414) 277-5591

INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:

LENGTH: 841 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Streptococcus pneumoniae

US-08-743-637B-34

Alignment Scores:

Pred. No.: 6,448-21 Length: 841
 Score: 304.50 Matches: 74
 Percent Similarity: 59.24% Conservative: 19
 Best Local Similarity: 47.13% Mismatches: 29
 Query Match: 7.31% Indels: 36
 DB: 2 Gaps: 3

US-09-765-271-56 (1-796) x US-08-743-637B-34 (1-841)

QY 663 AlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAspGlyTrp 682
 DB 836 GCGACTGTCACTACTATGTCGACATCCAAACGAAACGTCG-CATTGAGATAATGTTT 778
 QY 683 GlyAsnAlaSerGluHisValLeuGlyLysAspHisSerGluAspProAsnLysAsn 702

```

; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; US-08-526-840B-34

Alignment Scores:
Pred. No.: 6,44e-21 Length: 841
Score: 304.50 Matches: 74
Percent Similarity: 59.24% Conservatives: 19
Best Local Similarity: 47.13% Mismatches: 29
Query Match: 7.31% Indels: 36
DB: 3 Gaps: 3

US-09-765-271-56 (1-796) x US-08-526-840B-34 (1-841)

Qy 663 AlaThrIleLysTyrValGluHisProAspGluArgProHisSerAsnAspGlyTrp 682
Db 836 GCGACTGTCAAGTACTATGTGCAACATCCAAACGACGTCG-CATTTCAGATAATGGTTT 778
Qy 683 GlyAsnAlaSerGluHisValLeuGlyLysAspHisSerGluAspProAsnLysAsn 702
Db 777 GGTAACCTAGCACCCTGTTCAAGAAACAAATAATGTCAGCTGATCAATCAAACG 718
Qy 703 PhelysAlaAspGluGlu-----ProValGluGluThr----- 713
Db 717 GAAACCAACACGAGGAGAAACCTCAGACAGAAACCTCGAGGAGAAACCCCTCGAGAA 658
Qy 713 ----- 713
Db 657 GAGAAACCGCAAGCGAGAAACCGAGTCTCCAAACCAACAGAGAACCCAGAGAAAGAA 598
Qy 714 -----ProAlaGluProGluValProGluValGluThrGluLysValGlu 728
Db 597 TCACCAAGAAATCACCCAGAGAAATCAGAGAACCTCAGGTCGAGACTGAAGGTTAAGA 538
Qy 729 AlaGlnLeuLysGluAlaGluValLeuAlaLysValThrAspSerSerLeuLysAla 748
Db 537 ---AACTGAGAGAGCTGAGATTTACTTGGAAAAATCCAGAAATCCAATTATCAAGTCC 481
Qy 749 AsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsn 768
Db 480 AATGCCAAAGAGACTCTCAGAGGATTAATAATAATTTACTATTGTCACCCAGGACAAC 421
Qy 769 AsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer 785
Db 420 AATACTATTATGCGAGAGCTGAAAAACTATTGGCTTTATTAAGGAGAGT 370

RESULT 17
US-08-526-840B-34/c
; Sequence 34, Application US/08526840B
; Patent No. 6001564
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,840B
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/304,732
; FILING DATE: 12-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 841 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

```

```

; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; US-08-526-840B-34

Alignment Scores:
Pred. No.: 6,44e-21 Length: 841
Score: 304.50 Matches: 74
Percent Similarity: 59.24% Conservatives: 19
Best Local Similarity: 47.13% Mismatches: 29
Query Match: 7.31% Indels: 36
DB: 3 Gaps: 3

US-09-765-271-56 (1-796) x US-08-526-840B-34 (1-841)

Qy 663 AlaThrIleLysTyrValGluHisProAspGluArgProHisSerAsnAspGlyTrp 682
Db 836 GCGACTGTCAAGTACTATGTGCAACATCCAAACGACGTCG-CATTTCAGATAATGGTTT 778
Qy 683 GlyAsnAlaSerGluHisValLeuGlyLysAspHisSerGluAspProAsnLysAsn 702
Db 777 GGTAACCTAGCACCCTGTTCAAGAAACAAATAATGTCAGCTGATCAATCAAACG 718
Qy 703 PhelysAlaAspGluGlu-----ProValGluGluThr----- 713
Db 717 GAAACCAACACGAGGAGAAACCTCAGACAGAAACCTCGAGGAGAAACCCCTCGAGAA 658
Qy 713 ----- 713
Db 657 GAGAAACCGCAAGCGAGAAACCGAGTCTCCAAACCAACAGAGAACCCAGAGAAAGAA 598
Qy 714 -----ProAlaGluProGluValProGluValGluThrGluLysValGlu 728
Db 597 TCACCAAGAAATCACCCAGAGAAATCAGAGAACCTCAGGTCGAGACTGAAGGTTAAGA 538
Qy 729 AlaGlnLeuLysGluAlaGluValLeuAlaLysValThrAspSerSerLeuLysAla 748
Db 537 ---AACTGAGAGAGCTGAGATTTACTTGGAAAAATCCAGAAATCCAATTATCAAGTCC 481
Qy 749 AsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsn 768
Db 480 AATGCCAAAGAGACTCTCAGAGGATTAATAATAATTTACTATTGTCACCCAGGACAAC 421
Qy 769 AsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer 785
Db 420 AATACTATTATGCGAGAGCTGAAAAACTATTGGCTTTATTAAGGAGAGT 370

RESULT 18
US-08-956-171E-63
; Sequence 63, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E

```

```

; FILLING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILLING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILLING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8155 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 63:
US-08-956-171E-63

Alignment Scores:
Pred. No.: 6,39e-06 Length: 8155
Score: 170.00 Matches: 169
Percent Similarity: 35.36% Conservative: 139
Best Local Similarity: 19.40% Mismatches: 355
Query Match: 4.08% Indels: 208
Gaps: 38
DB: 4

US-09-765-271-56 (1-796) x US-08-956-171E-63 (1-8155)

QY 8 GlnAlaArgThrValLysGluAsnArgValSerTyrIleAspGlyLysGlnAlaThr 27
Db 4948 CAAGTGCACAACTAAGAAATCGATGCAAAAGCGAAATCGCTCAAAAAGCAAGTGA 5007
QY 28 GlnLysThr-----GluAsnLeuThrProAspGluValSerLysArg 41
Db 5008 COTAAAACAGCAATTGACGAAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5067
QY 42 GluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThrSer 61
Db 5068 GACAAAGTG-----GATCAAGCAGTAGTACTGCA 5097
QY 62 HisGlyAsp-----HisTyrHisTyrTyrTyrAsnGlyLysValProTyr 75
Db 5098 AACGCTGATATAGATAATGCTGCAGCAACCAATGATGCTGCTGCTGCTGCTGCTGCTG 5157
QY 76 AspAlaIleIleSerGluGluLeuMetLysAspProAsnTyrLysLysAspGlu 95
Db 5158 GAAGCTACAAUGCA-----GCCATTACCTGATGCAATGTTAAACCCAGCAGCAAA 5211
QY 96 AspIleValAsnGluValLysGlyTyrValIleLysValAspGlyLysTyrTyrVal 115
Db 5212 CAAGCAATTGCAGATAAAGTCAAGCTCAAGAAACAGCAATTTGATGCA 5259
QY 116 TyrLeuLysAspAlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGln 135
Db 5260 -----AATAACGGCTCAACCACTGAAGAAACAGCAGTGT 5295
QY 136 LysGlnGluHisSerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAla 153
Db 5296 AAACACACAGTTCAAACTGAAACCAACACAGCTGATGCGCAATAGATGACACATACA 5355
QY 154 -----ValAlaLeuAlaArg-----SerGlnGly 161
Db 5356 AATCGGAAAGTTGAAGCGCTAAAAAGCAGCAATTTGCTTAAATTAAGCAGTATTCACCA 5415
QY 162 ArgTyrThrThrAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThrGly 181
Db 5416 GCAACAACTAAAGAT-----AATGCGAAAGACCAATTCCTACGAAGCG 5463

```

```

QY 182 AspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeuSer 201
Db 5464 AATCAACGTAAACAGCA-----ATCGCTCAACGCAAGCAATTAAT 5505
QY 202 AlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSer 221
Db 5506 GCTGAAGAAATTCAGCGGCTAATGCG-----GACGTAGATAATGCT 5547
QY 222 ArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSerVal 241
Db 5548 GTGACA-----CAAGCAATAGCAACATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5598
QY 242 SerAsnProGlyThrThrAsnThrAsnThrSerAsn-----AsnSerAsnThrAsnSer 259
Db 5599 GACCAAGCGAAACGACAGGTGAAATAGTATTGATCAAGTAACACCAACAGTTAATAAA 5658
QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
Db 5659 AAAGCACTGCACGTAATGAATCAGCAATTTTAAATAAC-----AAATTGCAAGAG 5712
QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
Db 5713 ATTCAAGCTACGCCAGATGCAACAGTAAGCAAAACAGCAGCTGCTGCTGCTGCTGCTGCTG 5772
QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
Db 5773 ACTGAAATGCTAAAGCA-----AATCAA 5796
QY 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
Db 5797 GCCATTTTCAGCAGCAACTACTAAGCAAGCTGATGAAGCTAAAGCAAAATGAGCAAGCA 5856
QY 340 TrpValProAspSerArgProGlnProSerProGlnProThrProGluProSerPro 359
Db 5857 CGGATTATCGGTAAACCAAAAGTTGTGAAGAAACAGCGCTTAAGATCAATTTGAT 5916
QY 360 GlyProGlnProAla-----ProAsnLeuLysIleAspSer 371
Db 5917 CAATTACAAACGCAACCAAAATGTTATCAATATGATCAGAACGCTACACAGCAAGAA 5976
QY 372 AsnSerSerLeuValSerGlnLeuValArgLysValGlyGluGly----- 386
Db 5977 AAAGAACGAGCTATTCAACAATTTAGCAACAGCAGTTACAGACGCGGCAAAATAATATTACA 6036
QY 387 TyrValPheGluGlyLysGlyIleSerArgTyrValPheAlaLysAspLeuProSerGlu 406
Db 6037 GCTGCACTGATGATATGTTGTTAGTACAG-----GCGAAAGCGCTGGAAGAAT 6087
QY 407 ThrValLysAsnLeuGluSer-----LysLeuSerLysGlnGluSerValSer 422
Db 6088 TCAATTCAAGCAGCAGCAACCAACAGCGGTTAAATCAAAATGCTTAAATAATGATTTGAT 6147
QY 423 HisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyrAspLys 442
Db 6148 CAAGCTGTGACAACTCAAAATCAAGCAAT-----GATAAT 6183
QY 443 AlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeu-----Phe***AsnLysGlyArg 460
Db 6184 ACAACTGCTGTCTACAACTGAAGAGAAAAATGCAAGCAAAAGATTTAGTTTAAAGCTAAA 6243
QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
Db 6244 GAAAAAGCGTATCAAGATATC-----TTAAATGCAACAAACCAACTAAT 6285
QY 481 -----LysLysLysLeuValAspLeuLeuAlaPheLeuAlaPheLeu 495
Db 6286 GATGTTACGCAATTAAGATCAAGCTTGTGTTATTTCAAGCTATTACTGCAATACA 6345
QY 496 ThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluVal 515
Db 6346 ACAATTAAAGAT---GTTGCGAAAGATGAATTAAGCAACAAAGCAAAACGCAAAAGCG 6402
QY 516 ArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHis 535

```

```

Db 6403 CTTATTGCACAACTGCAGATCGGACTACT.....:GAAGAAAA 6441
Qy 536 AspilelleSerAspGluGlyAspAlaThrProHisMetGlyHisSerHisTrp 555
Db 6442 GAACAGCAATCAACAGATGAGCCACAAATTAACA-----CAGGTAATCAAAATATT 6495
Qy 556 IleGlyLysAspSerLeuSerAspLysGluLysValAlaGlnAlaThrLysGlu 575
Db 6496 GAAATGACACAGTCAATCGATGATTAACACTGCAGAAAGATAATGCAATT----- 6546
Qy 576 LysGlyLleLeuProSerProAspAlaSerValLysAlaAsnProThrGlyAspSer 595
Db 6547 CAAGCAATTTGACCAATTTCAAGCATCAACAGATGTTAAACGAATGCAAGCGCAATTG 6606
Qy 596 AlaAlaAlaIleTyAsnAtgValLysGlyGluLysArgLleProLeuValArgLeuPro 615
Db 6607 CTAACGAAATGCAAAATAAAATA-----ACTGAAATACTT 6642
Qy 616 TyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIlePro----- 632
Db 6643 AATAATAATGAGACTACTAATGAAGAAAAA---GGTAACGATATGACCGATTAGAGCA 6699
Qy 633 ---HisLysAspHisTyHisAsnIleLysPheAlaTrpPhe-----AspAspHisThr 649
Db 6700 GCATATGAAGAAGGTTTAAATAATTAATTAATGACGCACTACTACAGGTGATGTAATACT 6759
Qy 650 TyrlsAlaProAsnGlyTyThrLeuGluAspLeuPheAlaThrLleLysTyTrVal 669
Db 6760 GCTAAGATACAGCAGTACAAAAGTTTCAACACTTCATGCA----- 6801
Qy 670 GluHisProAspGluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisVal 689
Db 6802 ---AATCCTGTGAAGAAACCA----- 6819
Qy 690 LeuGlyLysLysAspHisSerGluAspProAsnLysAsnPheLysAlaAspGluGlu--- 708
Db 6820 GCAGGTAAAAAAGAAATTAGATCAAA-----GCTGCAGCTGATAGAAAAACA 6864
Qy 709 ProValGluThrPro-----AlaGluProGluValProGlnValGluThrGluLys 726
Db 6865 CAAATGACAAACACCAACCAATGTCATCAACAAGAAATTAATGATGCAAAACAGAA--- 6921
Qy 727 ValGluAlaGlnLeuLysGluAla-----GluVal 736
Db 6922 GTTGATACTGAATTAATTAATCAAGCGAAAAAATGTCATCAATCATCAACAATGATAT 6981
Qy 737 LeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGly 756
Db 6982 GTTGATAATGCAGTTAAAGAAAGAAAGCTAAATAATTAATGAGTTAAACATTTAGTGAG 7041
Qy 757 LeuArgAsnAsnLeuThrLeuGlnIleMetAsp----- 767
Db 7042 TACAAAAAGATGCTTTAGCTAAATTTGAAGATGTCATATTAATGCTAAAGTAAACGAGCG 7101
Qy 768 -----AsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeu 781
Db 7102 GATAACTCTAACGATCGACTTCAAGTGAATTTGCTGACGGAACAAAACTTGCTGAA 7161
Qy 782 LeuLysGlySerAsnProSerSerValSerLys 792
Db 7162 TTAATAACAACTCGCGATCAAAATGTTAATCAA 7194

```

RESULT 19

```

US-08-956-171E-310
; Sequence 310, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
;           Gil H. Choi
;           Patrick S. Dillon
;           Craig A. Rosen
;           Steven C. Barash

```

```

; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

```

COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text

```

CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997

```

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996

```

```

; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997

```

ATTORNEY/AGENT INFORMATION:

```

; NAME: Mark J. Hyman

```

```

; REGISTRATION NUMBER: 46,789

```

```

; REFERENCE/DOCKET NUMBER: PB248P1

```

TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439

```

INFORMATION FOR SEQ ID NO: 310:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 12173 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

```

```

; SEQUENCE DESCRIPTION: SEQ ID NO: 310:
US-08-956-171E-310

```

Alignment Scores:

```

Pred. No.: 1-54e-05 Length: 12173
Score: 169.00 Matches: 179
Percent Similarity: 34.99% Conservative: 138
Best Local Similarity: 19.76% Mismatches: 343
Query Match: 4.06% Indels: 246
DB: 4 Gaps: 46

```

```

US-09-765-271-56 (1-796) x US-08-956-171E-310 (1-12173)

```

```

Qy 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeu---ThrProAspGluValSer 39
Db 9540 ATTACTTCTCAACATCAAGCAGCAGCAAGAAATACAAATACTTTCAGATAAAATCTCG 9599
Qy 40 LysArgGluGlyLleAsnAlaGluGlnIle---ValIleLysIleThrAspGlnGlyTyr 58
Db 9600 GAAATCAAAATAAATAATGCAATTAAGATCTCTGCATTCGATTAAGATAAAATCAACACAA 9659
Qy 59 ValThrSerHisGlyAspHisTyHisTyAsnGlyLysValProTyTrAspAlaIle 78
Db 9660 CCGTCTACCAACCA-----GCACACACTGCGGAAAACTATCTCTGCA--- 9701
Qy 79 IleSerGluGluLeuLeu-----MetLysAspProAsnTyTrLysLeuLysAspGlu 95
Db 9702 ---GCGGATGAATCACTTAAGATGCAATTAAGATCTCTGCATTCGATTAAGATAAAATCAACACAA 9758
Qy 96 AspIleValAsnGluValLysGlyGlyTyrValIle-----LysValAspGlyLys 112
Db 9759 GATATAGTGTCCAGAGCAACAGTCAATTTCCAGTTTATTAGATAAAACAAATGAACGCAG 9818
Qy 113 TyTrTyValTy-----LeuLysAspAlaAlaHisAlaAspAsnValArgThrLysGlu 130
Db 9819 TACTATCACTTTTTCAGCATCAAGATCCAGCATGTGTATTACACTAAAGAAAGCA 9878

```

Qy 131 GluUle-----Asn 133
 Db 9879 GAAGTTGAATTAGACATCAATACTGCTTCAACATGGGAAGATTGGAAGCTTATGAAAAC 9938
 Qy 134 ArgGlnLys-----GlnGluHisSerGln 141
 Db 9939 AATCAAAAATGGCCAGTCAGACTTGTATCATATAGTCTCTGTCACCAAGAACCATGCTAT 9998
 Qy 142 HisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGlnGly 161
 Db 9999 ATTCCGA-----TTCCAGATTTCAGATGCCACACAAGAATTG--AAAATTGTTTCT 10046
 Qy 162 ArgTyrThrThrAspAspGlyTyr-----169
 Db 10047 TCGACTCAAAATTGATGGGGAAGAAACAAATATTGATTGTTTACTAAATTAGTATTGCT 10106
 Qy 170 -----IlePheAsnAlaSerAspIleIleGlu--AspThrGlyAspAlaTyrIleVal 186
 Db 10107 AAACCTATTATACGATCCTTCACTTGTAAAATCAGATACAAATGATGTCAGTAGTAACG 10166
 Qy 187 ProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeuSerAlaSerGluLeuAla 206
 Db 10167 -----AATGATCAATCAAGTTTCAGTCGCAAGTAAATCAAAACACCAATACATCTAAT 10220
 Qy 207 AlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSerArgThrTyrArgArg 226
 Db 10221 CAAAATATATCAAGATCAACATCTAATCAATCAACCGCAGGCACCAATATGAGT 10280
 Qy 227 GlnAsnSerAspAsnThrSerArgThrAsnTrpValProSerValSerAsnProGly--- 245
 Db 10281 CAACCTGCACACCAAAATCGTCAAGAAATGAGATCAAGCTCAAGCGCAACCACTCAT 10340
 Qy 246 ThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAlaSerGlnSerAsn 265
 Db 10341 GAAACAAATTTCTAATGGTAATCTAATCAATCAATCAATCAATCAATCAATCAATCAAT 10400
 Qy 266 AspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSerGlnArgHisValGlu 285
 Db 10401 -----GTTAATCAACAGTATCCACCA 10421
 Qy 286 SerAspGlyLeuValPheAsp-----ProAlaGlnIleThrSerArgThrAla 301
 Db 10422 GCAGATGAATCACTACAGATGCAATTAATAAACCAGCTATCATCGATAAAGAACATACA 10481
 Qy 302 ArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMetSer 321
 Db 10482 -----OCTGATATATGGCGACCAATTTGATTTTCAAAATGAAAAT 10520
 Qy 322 GluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrpVal 341
 Db 10521 GATAAAGGTGAAGA-----CAGTTCTATCATTTATGCT 10553
 Qy 342 ProAspSerArgProGluGlnProSerProGlnProThrProGluProSer----- 358
 Db 10554 -----AGTACTGTTGAAACCAACTGTCAATT 10580
 Qy 359 -----ProGlyProGlnProAlaProAsnLeuLys-----368
 Db 10581 TTTACAAAACAGGACCAATAATTAATTAAGTTTAAAGACAGCTTCAACATGGGAAGAAA 10640
 Qy 369 -----IleAspSerAsnSerSerLeuValSerGlnLeuValArgLysValGlyGlu 385
 Db 10641 TTTGAAGTTTATGAAGTGACAAAAGTTACCAGTCGAATTAGTATCATATGATTTCTGAT 10700
 Qy 386 -----GlyTyrVal---PheGluGluLysGlyIleSerArgTyrValPheAlaLys 401
 Db 10701 AAAGATTATGCTATATTCGTTTCCAGTATCTAATGTTGACGAGAGAAGCTT----- 10751
 Qy 402 AspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSerVal 421
 Db 10752 -----AAAATTGTCATCTATTGATATGTTGTCAGNACATCCATGAGACTAT 10799

Qy 422 SerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyrAsp 441
 Db 10800 GATTATACGCTAATGGTCTTTTGCACAGCTATTACTAATAACCCAGACGACTGTGAT 10859
 Qy 442 ---LysAlaTyrAsn-----LeuLeuThrGluAlaHisLysAlaLeuPhe**Asn 457
 Db 10860 GAAGAAACATACAAATTTACAAAATATTAGTCTCCGTATCACAAGCT----- 10907
 Qy 458 LysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGlu 477
 Db 10908 AAAAGCTTAGAAGACACAAAGTTTATGAATTAGAAAATTTACAAGAGAAATTCGCCAGAAAA 10967
 Qy 478 ---SerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla----- 493
 Db 10968 TATAAGCGGAATATAAAAAGAAATTTAGTCAAACTAGAGTAGAGTTAGCTGATCAAGTT 11027
 Qy 494 -----ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThr 511
 Db 11028 AATCAGCAGTGACGGAATTTGAAAATGTTACACTACAAATGATCAATTAACAGATTTA 11087
 Qy 512 GluAspGluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSer-----AspGly 529
 Db 11088 CAAGAAGCGCATTTTGTGTTTGAAGTGAAGAAATAGTCAAGTCAAGTTATGACGCGC 11147
 Qy 530 TyrIlePheAspGluHisAspIleIleSerAsp-----GluGlyAspAlaTyrValThr 547
 Db 11148 TTTGTT-----GAACATCCATCTATACACCACTTTAAATGTTCAAAAATATGATG 11201
 Qy 548 ---ProHisMetGlyHisSerHisTrpIleGlyLysAspSerLeuSerAspLysGluLys 566
 Db 11202 ATGAAAACAAAGGATGACAGTTACTGG-----AAAGATTAAATTTGTAAGGTAACGT 11255
 Qy 567 ValAlaAlaGlnAlaTyrThrLysGluLysGly-----IleLeuProProSer 582
 Db 11256 GTCACTACTGTTTCTAAAGATCCTAAATAATTTCTAGAACGCTGATTTTCCATATATA 11315
 Qy 583 ProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyrAsnArg 602
 Db 11316 CTGAC-----AAAGCAGTTTACAACTGCG 11339
 Qy 603 ValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHisThrVal 622
 Db 11340 ATT-----GTTAAAGTGTGTGCAACATTTGTTATGAAGTCAATATCATGTC 11390
 Qy 623 GluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIle----- 640
 Db 11391 AGAATTATAATCAGGATATCAATACAAAAGATGATGATACATCACAAAATAACACGAGT 11450
 Qy 641 -----LysPheAlaTrpPheAsp-----Asp 647
 Db 11451 GAACGCTAAATGTACAAACAGGACAAAGAGTAAAGTTGCTGATACAGATGATGCTGAA 11510
 Qy 648 HisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyr 667
 Db 11511 AATACGACACTGCAACAAATCCTAAAGATGCTGTGATAAAGCAGATGTGATAGAA--- 11567
 Qy 668 TyrValGluHisProAspGluArgProHisSerAsn-----AspGlyTrpGlyAsn 684
 Db 11568 -----CCAGAGTCTGAGTGTAAAGATGCTGATATAAT 11603
 Qy 685 AlaSerGluHisValLeuGlyLysAspHisSerGluAspProAsnLysAsnPhelLys 704
 Db 11604 ATTGATAAAGTGTCAACATGATGTTGATCATTTATTCGATATGTCGGAATATAATCATC 11663
 Qy 705 AlaAspGluGluProValGluGluThrProAlaGluProGluValProGlnValGluThr 724
 Db 11664 TTTCAT----- 11669
 Qy 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer 744
 Db 11670 -----AAATATGATTAAAGAAATGGTACTCTCAAAATGCCAAAGATCTGATAGA 11720
 Qy 745 SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeu--- 763

Qy	431	-----AsnValAlaProArgAspGlnGlu-----	438
Db	5167	AAGCGTCAAAATCAATGCAAGCCCAATGCTACAGAAAGAAAAACAAGATCAATCAAA	5226
Qy	439	-----PheTyrAspLysAlaTyrAsnLeu-----	447
Db	5227	AGATTGAATGGTAAAGAGAGATGAAGTTAAGATCTAATAAATCAAGATAGACGTGACAAT	5286
Qy	448	---ThrGluAlaHisLys-----AlaLeuPhe***AsnLys	458
Db	5287	GAAGTTGAACAGACACAAATAATTGGACTTCAAGAATTAGAAACGATTATCATGCTAATCCA	5346
Qy	459	GlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeu-----	474
Db	5347	ACTAGAAAATCTGAT-----GGCTCCGAAGAGTTACAAACTAAATTTATTTCCAAACA	5400
Qy	475	-----AsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeu	489
Db	5401	GAGTTAATTAAATAACAAGATGCACTAATGAAGAAAA-----GATGAAGCCAAA	5454
Qy	490	AlaPheLeu-----AlaProIleThrHisProGluArgLeuGlyLysPro	504
Db	5455	CGACTTCTTGAGATTAGTAAATAATAAACTATAACAATAATCAATCAA-----CGCAAACT	5511
Qy	505	AsnSerGlnIleGluTyrThrGluAspGlu-----ValArgIleAlaGlnLeu-----	520
Db	5512	AATAATCAAGTTGATAATGCTAAAGATPAACGGCATGAATGAGATGCTACCAATAATCCA	5571
Qy	521	-----AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHis	535
Db	5572	GCAACAACAATAAACAAGATGCAAAACCGCTATTGATAAAAAGCTGAGCAACAAGTT	5631
Qy	536	AspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyr	555
Db	5632	ACANTCATCAATGCTAAACAAGATGCA-----	5658
Qy	556	IleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGlu	575
Db	5659	-----ACAGATGAAGAAAAAGCAGAGCTAGAAAGCTGGTGTGAAAAA	5700
Qy	576	LysGlyIleLeuProSerProAspAlaAspValLysAlaAsnProThrGlyAspSer	595
Db	5701	GCGAAATTT-----GAAGCCAAATCTAATATTACAAATAGTGAT	5739
Qy	596	AlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArgLeuPro	615
Db	5740	ACTGAA-----AGGGAAGTCAATGGTGTCTAAAACCAATGGGTAGAAAAATAAAC	5790
Qy	616	TyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAsp	635
Db	5791	AATATTCAACCATCAACTCAAACTAAACAAATGCTAAGCAAGAAAATAATGCAAGCT	5850
Qy	636	HisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro-----	653
Db	5851	CAAGAACAAATTAATCCAAATTAATAACACGCTGATGCAACCGAAGAAAAGCAAGAG	5910
Qy	654	-----AsnGlyTyr	656
Db	5911	GCAACAAATAGATCAATGCTGGATTAGCACACAGCAATACAAAATAATAATTAATGACAT	5970
Qy	657	ThrLeuGluAspLeu-----PheAlaThrIleLysTyrTyrVal	669
Db	5971	AGTACTCAAGAGTAAGTAATGAATCTAAACAAATAATAGTATTGCTACAACTCAGAGTGTCACAA	6030
Qy	670	GluHisProAspGluArgPro-----HisSerAsnAspGlyTyrGlyAsnAlaSer	686
Db	6031	CCCAATGTGATCAAAAAACCGACTGCTATAATAGTTTGACTCAAGAGCTTAATAATCAA	6090
Qy	687	GluHisValLeuGlyLysLysAspHisSerGluAspProAsnLysAsnPhe-----	703
Db	6091	AAGACGTTAATAGGTAATGATGGTAATGCTACTGATGATGAAAAAGAGGCTGCAAGCAA	6150

QY	704	-----LysAlaaspGlu-----GluProValGluGluThrProAlaGluPro	717
Db	6151	TTAGTGACCCAAAATTTAAATGAACAATAATTCAAAATAATTCATGAAGTACACAAGATAAT	6210
QY	718	GluValProGlnValGluThrGluLysValGluAla	730
Db	6211	CAAGTTGATAACGTAAAAGCACAGCTATCACTGCAATTAATGATTAAATGATGCAAAATGCA	6270
QY	731	LeuLysGluAlaGluValLeu	745
Db	6271	CATAAAGACAAAGATGCATTAATATTTTGACTAATCTAGCTGAAAGTAAAAAATCAGAT	6330
QY	746	LeuLysAlaAsn-----AlaThr	751
Db	6331	ATAGAGCCCAATCAAGATGCACTACTGAAGAGAAAAATACGCAATACAATCTATAGAT	6390
QY	752	GluThrLeuAlaGlyLeuArgAsnLeu-----ThrLeuGlnIleMetAsp	767
Db	6391	GATACGTTAGCACAAAGCACGCTAAACAATTAATGGTGCATAATCAAAATACAAATGCGTTAGTGAT	6450
QY	768	AsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLeuLysGlySerAsnPro	787
Db	6451	GAGAAAT---TTAGAAGATGGTAACGAAGAAGTTTCAACGTAATGTTGTTCAACTCAAACT	6507
QY	788	SerSerValSerLysGluLysIle	795
Db	6508	AAAACACAAAGCTAAAGCAGACATT	6531
RESULT 21			
US-09-134-001C-322			
; Sequence 322, Application US/09134001C			
; Patent No. 6380370			
; GENERAL INFORMATION:			
; APPLICANT: Lynn Doucette-Stamm et al			
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHY			
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS			
; FILE REFERENCE: GTC-007			
; CURRENT APPLICATION NUMBER: US/09/134,001C			
; CURRENT FILING DATE: 1998-08-13			
; PRIOR APPLICATION NUMBER: US 60/064,964			
; PRIOR FILING DATE: 1997-11-08			
; PRIOR APPLICATION NUMBER: US 60/055,779			
; PRIOR FILING DATE: 1997-08-14			
; NUMBER OF SEQ ID NOS: 5674			
; SEQ ID NO 322			
; LENGTH: 30549			
; TYPE: DNA			
; ORGANISM: Staphylococcus epidermidis			
US-09-134-001C-322			
Alignment Scores:			
Pred. No.:			
Score: 0.000676 Length: 30549			
Percent Similarity: 159.00 Matches: 189			
Best Local Similarity: 34.88% Conservative: 141			
Query Match: 19.98% Mismatches: 330			
DB: Indels: 288			
Gaps: 49			
US-09-765-271-56 (1-796) x US-09-134-001C-322 (1-30549)			
QY	6	LeuTyrglnAlaArgThrValLysGluAsnAsnArgValSerTyrlleAspGlyLysGln	25
Db	7378	GTATTTTCGGCAATACAACTAACTAACTCTCAAAATTACA---ATAACTCTTAAAGCA	7430
QY	26	AlaThrGlnLysThrGluAsnLeuThrProaspGluValSerLysArgGluGlyIleAsn	45
Db	7435	GGTCAGGGGTAACTCTGAAAAACAATAATCCT	7466
QY	46	AlaGluGlnIleValIleLysIleThrAspGlnGlyTyrrValThrSerHisGlyAspHis	65
Db	7465	-----ACTGTAACTCAAGCACCCTGGCAACATACCTTTAACA-----	7500
QY	66	TyrHisTyrrAsnGlyLysValProTyrrAspAlaIleSerGluLeuLeuMet	85

Db	7501	:	-----ATCAATGAA-----	:	-----ATTGTT	7515
Qy	86	lysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGluValLys-	:	-----	-----	102
Db	7516	AAAGAACAGGTCAAAATGTGACTAAATGATGATATTAATATGCGTTCAAGTGCCAAAT	:			7515
Qy	103	-----	-----	-----GlyGlyTyr	-----	105
Db	7576	AAAAATAGAGTTGCGATTAAACAAGRAACGCTCTTCCAACAATTTAGCTGGTGAGT	:	-----	-----	7635
Qy	106	ValllleLysValAspGlyLysTyrTyrValTyrLeuLysAspAlaHis-	:		-----	122
Db	7636	ACATCACAATAT-CCAGTAGTTATTATTATACAGTGTGCAAGCTTCTGAAGAAGCT	:		-----	7689
Qy	123	AlaAspAsnValArgThrLys-----GluGluIleAsn-----ArgGlnLysGln	:		-----	137
Db	7690	ACTGAGACTGTTAGACTAAAGTTAATAAACCAGTAATTAATCAATGCTCGTCGCACCTA	:		-----	7749
Qy	138	GluHisSerGlnHisArgGluGlyGlyThrPro-----ArgAsn-----AspGlyVala	:		-----	153
Db	7750	GATGAAGAAATTTAGTAAGAAGAACAAACACCATCAAGTATCAGAACTTTGATCAAGCT	:		-----	7809
Qy	154	ValAlaLeuAlaArgSerGlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAla	:		-----	173
Db	7810	ATGAATCGTCTCAATCACAATTTAATACAGCTAATAAGTGTAT-----GCT	:		-----	7854
Qy	174	SerAspIleIleGluAspThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHis	:		-----	193
Db	7855	GACCAAGTTATA-----GGCACAGAAATTCGAACACCTCAACAGTAAT-----TCA	:		-----	7902
Qy	194	TyrIleProLysAsnGluLeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSer	:		-----	213
Db	7903	GCTTTATCTAAAGTTCAAGCGGCACAAAATAAATAAATGAAGCTAAAGCAATTTATCAA	:		-----	7962
Qy	214	GlyArgGlyAsnLeuSerAsnSerArgTyrTyrArgArgGlnAsnSerAspAsnThrSer	:		-----	233
Db	7963	AACAAGCTCGAT-----AATAGTCAACTGTGAGAGCAAAAGAACAAATTACAACAATCG	:		-----	8016
Qy	234	ArgThrAsnTrpValProSerValSerAsnProGlyThrThrAsnThrAsnThrSerAsn	:		-----	253
Db	8017	ATTCAA-----CCAGCGCTTCAACTGATGTTGATGACTCAAGATAGCACAGGAAC	:		-----	8067
Qy	254	AsnSerAsnThrAsnSerGlnAlaSerGlnSer-----	:		-----	264
Db	8068	TACAAAATTAACGCCMAGCAGCTGAACAGCAATACAAATGCAATAGCGTTATAAT	:		-----	8127
Qy	265	-----AsnAspIleAspSerLeuLeuLysGlnLeuTyr	:		-----	275
Db	8128	AATGGAGATGCAACATCCCAACAAATTAATGATGCTAAAAACACAGTTGAACAGGCACAG	:		-----	8187
Qy	276	Lys-----LeuProLeuSerGlnArgHisValGluSer---	:		-----	286
Db	8188	AGAGATTGTTGAAGCTAAAGCAACTACGTGCTGATGAATCAAGTTACAAAGCGCT	:		-----	8247
Qy	287	-----AspGlyLeuValPheAsp-----ProAlaGlnIle-----	:		-----	296
Db	8248	TATGATACGTTAAATAGAGATGTTTTTAAACAATGATAAAAACCCAGCATCTGTAAAGCGC	:		-----	8307
Qy	297	-----ThrSer	:		-----	298
Db	8308	TATAATGAAGCCATTTCAAAATATTAGAAAAAGAAATTAGATACAGCTAAAGCGATCCAAGT	:		-----	8367
Qy	299	ArgThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSer	:		-----	318
Db	8368	AGTACTTGGAAACACC-----AATCCT	:		-----	8391
Qy	319	GlnMetSerGluLeuGluGlnAlaArgIleIleProLeuArgTyrArgSerAsn	:		-----	338
Db	8392	TCCGTTTGAACAGTTAGAGCGCTTTAAATAAATAAATACTGTTCAACCTAAAGTGAAT	:		-----	8451
Qy	339	HisTrpValProAspSerArgProGluGlnProGlnProGlnProGlnProSer	:		-----	358

8452	DB	CAAGCAATGCTTTACTTCAACCCAAAAAATAATTACAGAA---CTTGTCAAGACTAAA	8508
359	QY	ProGlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGln	378
8509	DB	AAACGTTTTCAAGACGCTGTAAT---GACATACCTCAACACACAGGTATGACACAACNA	8565
379	QY	LeuValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrVal	398
8566	DB	ACAATTTAATAATTAATATGACAAACACGCTGAAGCTGAAAGACACATTACATCTGCACAA	8625
399	QY	PheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGln	418
8626	DB	AGATGATTGATAATGGGGATGCTACAACTCAAGAAATTACTCTGAAAATCTTAAAGTA	8685
419	QY	GluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGlu	438
8686	DB	GAGCAAGCAATGCAAGCTTTAACTTAATGCTTAAAGTAATCTGAGA---GCTGATAAGAT	8742
439	QY	PheTyrAspLysAlaTyrAsnLeuLeuThr-----	448
8743	DB	GAGTTACAGACTGCATATAACAATTAATTGAGAACGCTATCTACCAATGTAATAAACCG	8802
449	QY	-----GluAlaHisLysAlaLeuPhe***AsnLys-----	458
8803	DB	CGGAGTATAGCTCAATACGAAACAGCCAAAGCCAGACATACAAATCAAAATTAATGATGCT	8862
459	QY	-----GlyArgAsnSerAsp-----	465
8863	DB	AAAAATGAAGCGGCGAAATTTAGTGAATGATAATCCACAGTATCACAGTAACTCAA	8922
466	QY	AlaLeuAsp-----LysLeuLeuGluArgLeu-----	474
8923	DB	GCATTGAACAAAATCAAAGCTATTCAACCAAAATTAACAGAGCTATCAACATGCTTCAA	8982
475	QY	-----AsnAspGluSerThrAsnLysLysLeuValAspAspLeuLeuAla	490
8983	DB	AACAAAGAAATATACAGAAATGATCAATGCTTAAAAACAGACTTTGAAAATCAGTAAAT	9042
491	QY	PheLeuAlaProIleThrHis-----ProGluArgLeuGlyLysProAsnSerGln	507
9043	DB	GATACAGATCCA---ACACAGGTATGACTCAGAAACAAATTAATTAACACAGCGTAAA	9099
508	QY	IleGluTyrThrGluAspGluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSer	527
9100	DB	AAGCGAAGACTCAAAATGAATACAAAAAGCG-----	9132
528	QY	AspGlyTyrIlePheAspGluHisAspIleIleSerAspGluGlyAspAlaTyrValThr	547
9133	DB	-----AACATGATTAATAATGAGAGATGCT-----	9159
548	QY	ProHisMetGlyHisSerHisTrpIleGlyLysAspSerLeuSerAspLysGluLysVal	567
9160	DB	-----ACTGCTCAAGATATTTCTCTCGAAAATCTTAAAGTA	9195
568	QY	-----AlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuProProSerProAspAla	585
9196	DB	GAGCAAGTATTACAAGCATTAACAAAATGCTAAGAAT-----	9231
586	QY	AspValLysAlaAsnProThrGlyAspSerAlaAlaAlaIleTyrAsnArgValLysGly	605
9232	DB	GACTTAAGAGCTGATAAAAAGAAATTAACAGACTGCA-----TACAAATAAACTTA--TAC	9283
606	QY	GluLysArgIleProLeuValArgLeuProTyrMetValGluHisThrVal-----	622
9284	DB	AAATGTTAATACCAATGCTAATA---AACCATCTAGTATTCAAAACATTAAGTCTGCAA	9340
623	QY	-----GluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIleLys	641
9341	DB	GACGAAATATCGAAAAACCAAT---ATAATACCGCT-AAAAATGAAGCACATAATGTTCTT	9396
642	QY	PheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeu	661
9397	DB	-----GAAATTAACAAACCTACTGTAAAT-----GCAGTAAAGAT---	9432

```

Qy 662 PheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAspGly 681
Db 9433 ---GCTTACGTAAGATAAATGCAATTCACACGAGGTTTACA----- 9471
Qy 682 TrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsnLys 701
Db 9472 -----AAAGCTATTATATCTCAAGATTAAGAGATAATAGCGAACTTGTAGAGCA 9525
Qy 702 AsnPhelysAlaAspGlu-----GluProValGluGluThrProAlaGluPro 717
Db 9526 AAGAAATAATAGTCAAGCGATTAAATAGTCAACCATCACTAAATGGTATGACCTCAAGAA 9585
Qy 718 GluValProGlnValGluThrGluLysValGluAlaGln----- 730
Db 9586 TCTATTATAATATACACAAACAAACGTAAGAGAACACCAATTTAACG-----GCAGAT 9756
Qy 731 -----LeuLysGluAlaGluValLeuLeuAlaLysValThrAspSerSerLeuLysAla 748
Db 9646 ACTATTATATATGGGATGCATCTATTGAACAAATAACAGAAATAAATAATTCGAGTT 9705
Qy 749 ---AsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAsp 767
Db 9706 GAAGAGGCACTAATGCCTTAAACGAGCAAAACCAACATTTAACG-----GCAGAT 9756
Qy 768 AsnAsnSerIleMetAlaGluAlaGluLysLeuAlaLeuLeuLysGlySerAsn--- 786
Db 9757 ACAACTTCTTAAACAACTGAAGTACGGAATTTAAGTAGAGAGGCGCACAAACACAAA 9816
Qy 787 ---ProSerSerValSer 791
Db 9817 AAGCTAGAGGTAGT 9834

```

RESULT 22

US-08-956-171E-59
 ; Sequence 59, Application US/08956171E
 ; Patent No. 6593114

GENERAL INFORMATION:

APPLICANT: Charles Kunsch
 Gil H. Choi
 Patrick S. Dillon
 Craig A. Rosen
 Steven C. Barash
 Michael R. Fannon

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
 NUMBER OF SEQUENCES: 5256
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/956,171E
 FILING DATE: 20-Oct-1997

PRIOR APPLICATION DATA:

CLASSIFICATION: <Unknown>
 APPLICATION NUMBER: 60/009,861
 FILING DATE: January 5, 1996
 APPLICATION NUMBER: 08/781,986
 FILING DATE: January 3, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Mark J. Hyman
 REGISTRATION NUMBER: 46,789
 REFERENCE/DOCKET NUMBER: PE248P1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (240) 314-1224

TELEFAX: (301) 309-8439

INFORMATION FOR SEQ ID NO: 59:

SEQUENCE CHARACTERISTICS:

LENGTH: 31096 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 59:

US-08-956-171E-59

Alignment Scores:

Pred. No.: 0.00309 Length: 31096
 Score: 152.50 Matches: 171
 Percent Similarity: 32.80% Conservatives: 137
 Best Local Similarity: 18.21% Mismatches: 382
 Query Match: 3.66% Indels: 249
 Gaps: 39

US-09-765-271-56 (1-796) x US-08-956-171E-59 (1-31096)

```

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsn----- 15
Db 17213 TCACAGCTACAACTTAGCTGGTCTTGAATCTGTTAAACAAATGCGAATAGTTAGAT 17272
Qy 16 -----AsnArgValSerTyrIleAspGlyLysGlnAlaThrGlnLysThrGlu 31
Db 17273 GGTGCTATGGCTAACTTACAAACGGCTATCAACGATAAGTCAGGAACATTAGCGAGCCAA 17332
Qy 32 AsnLeuThrProAspGluValSerLysArgGluGlyIleAsn-----Ala 46
Db 17333 AACTTCTTTGGATGCTGATGAGCAAAACGTAATGTCATCAATCAAGCTGTATCAGCAGCC 17392
Qy 47 GluGlnIleValIleLysIleThrAspGlnGlyTyrValThrSerHisGlyAspHisTyr 66
Db 17393 GAAACCATTTTAAATAAACAACACTGA----- 17419
Qy 67 HisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSerGluGluLeuLeuMetLys 86
Db 17420 -----CCGAATACAGCAAAACAGCAGCTCGAACACAGCCTTAAT 17458
Qy 87 AspProAsn-----TyrLysLeuLysAspGluAspIleValAsnGluValLysGly 103
Db 17459 AATGTTAATAATGCGAAACATGCATTAATGTTAGTACGAAACTTAAACAAATCGGAACAA 17518
Qy 104 GlyTyrValIleLysValAspGlyLysTyrTyrVal-----TyrLeuLysAspAlaAla 121
Db 17519 GCAGCGATTACAGCAATCAATCGCATCTGATTTAAATCAAAACAAAGATGCATTAA 17578
Qy 122 HisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSerGln 141
Db 17579 AAAGCACAAAGCTAATGGTGCTCAACGGCTATCTAATGCAAGATGTACAGCAACATCGG 17638
Qy 142 -----HisArgGluGlyThrProArgAsn-----AspGlyAlaValAla 155
Db 17639 ACTGAACGACCGCAATGGGCATTTAAACATGCCATCGCAGATAAGCAATACG 17698
Qy 156 LeuAlaArgSer-----GlnGlyArgTyrThrThrAsp 166
Db 17699 TTAGCAAGCAGTAAATATGTTAATGCCGATAGCCTTAAACAAATAATCTTACACAACTAAA 17758
Qy 167 AspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThrGlyAspAlaTyrIleVal 186
Db 17759 -----GTTACCAATGCTGAACATATTTAGCGGTACGCCAACCGTTGTAGACAA 17809
Qy 187 ProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeuSerAlaSerGluLeu--- 205
Db 17810 CCT-----TCAGAAAGTAACA 17824
Qy 206 AlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSerArgThrTyrArg 225
Db 17825 GCTGCAGCTAATCAAGTAAACAGCGGAAACAGAAATTAATGTCGCGAAGATACGTT 17884
Qy 226 -----ArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSerValSerAsn 243

```

Db 17885 GAAGCAAAACAAACGCC-----AATACGCTATT 17914
QY 244 ProGlyThrThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAlaSerGln 263
Db 17915 GATGATTAACACAAATTAATTAACCTCAAAAGCTAAATTAAGAACAAGTGGGCA 17974
QY 264 SerAsnAspIleAspSerLeuLeuLysGlnLeuTyLysLeuProLeuSerGlnArgHis 283
Db 17975 GCCAATAGATTAGAGCGTA-----CAAACT 18001
QY 284 ValGluSerAspGlyLeuValPheAspProLalaGln-----IleThrSerArgThrAla 301
Db 18002 GTTCAAAACAATGACAGCACTTGAACAATCAATGAAGGCTTAAGAGATAGTATTGCT 18061
QY 302 ArgGlyValAlaValProHisGlyAspHisTyHisPheIleProTySerGlnMetSer 321
Db 18062 AACGAACAACAGCTCAAAACAAGTCAAAACTATATACAGACGCAAGTCCGAATACCAATCA 18121
QY 322 GluLeuGluGluArgIleAlaArgIleProLeuArgTyArgSerAsnHisTrpVal 341
Db 18122 ACATATAATAGCGTGTCTCAAAATGCGAAGGTATCATTAATCAAACTAACATCCGACT 18181
QY 342 ProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGlyPro 361
Db 18182 ATGGATACTAGTGGGATTTACCCCAAGCTACAAACAGTGAATATGCTTAAATAATGGTTA 18241
QY 362 GlnProAlaProAsnLeuLysIleAspSerAsnSer----- 373
Db 18242 AACGGTCTGAATACTTAAGAAATGCAAAACACTGCTAAGCAAAACTTAAATACATTA 18301
QY 374 -----SerLeuValSerGlnLeuValArgLysValGlyGlu 385
Db 18302 TCACACTTAACAATAACCAAAATCTGCCATCTCATCAAAATGATCGGAGTCAT 18361
QY 386 GlyTyValPheGluGluLysGlyIleSerArgTyValPheAlaLysAspLeuProSer 405
Db 18362 -----GTGAGTGAGGTAACCTGCTACTAATAAATGCGCAACT 18397
QY 406 Glu-----ThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSerValSer 422
Db 18398 GAGTGAATACGCAATGGGTAACTTGGCAACAGCTATCCATGATCAAAACACAGTTAAA 18457
QY 423 HisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyAspLys 442
Db 18458 CAAAGTGTTAAATTACTGATGCAGATAAGCTAAAGCTGAT----- 18499
QY 443 AlaTyLys-----AsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArgAsn 461
Db 18500 GCGTATACAAATGCGGTAAAGCAGAGCTGAAGCAATCTGAATAAAACGCAAGGTGCAAT 18559
QY 462 SerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsn----- 475
Db 18560 ACCTCTAAACAGATGTTGAAGCGGTATTCAAAATGTTTCAAGTGCTAAAAATGCAATTG 18619
QY 476 -----AspGluSer-----ThrAsnLysGluLysLeuValAspAspLeuAlaPheLeu 492
Db 18620 AATGGTGATCAAAACCTTACAAATGCAAGATGCGAGTAAATTAATTAATAACTTA 18679
QY 493 AlaProIleThrHisProGluArgLeuGlyLysProAsn----- 505
Db 18680 ACCTCAATTAATAATGCAAAAACGCTGACTTAACAACTAAATAATGATCAAGCAACAAT 18739
QY 506 -----SerGlnIleGluTyThrGluAspGluValArgIleAlaGln----- 519
Db 18740 GTAGCTGGTGTGAAGCTGTATCTAATACGAGTACACAAATGAATACAGCGATGGCTAAC 18799
QY 520 -----LeuAlaAspLysTyThrThr-----SerAspGlyTyThrIlePheAsp 533
Db 18800 TTGCAAAATGGTATTATGATAAAACAATACACTAGCAAGTGAATACTAT----- 18850
QY 534 GluHisAspIleLeuSerAspGluGlyAspAlaTyValThrProHisMetGlyHisSer 553
|||||

Db 18851 ---CATGATGCTGATTACAGATAAGAAAACCTGCTTATCTCAAGCGTTACGAACGAGAA 18907
QY 554 HisTrpIleGlyLysAspSerLeuSerAspLysGlyLysValAlaAlaGln----- 570
Db 18908 AATATTTTAAATAAATAATAGTATCAATTTAGACAAAACCTGCGTTGAAACGCGTTG 18967
QY 571 ---AlaTyThrLysGluLysGlyLysLeuProSer-----ProAspAlaAsp 586
Db 18968 TCACAAGTGTGTAATGCGAAAGTGCCTTAATGTAACTAATATTTAGAGCAAGCTAAA 19027
QY 587 ValLysAlaAsnProThrGlyAspSerAlaAlaIleTyArgAsnArgValLysGlyGlu 606
Db 19028 TCAAATGCAACACACTACT-----ATAACGCGACTT 19057
QY 607 LysArgIleProLeuValArgLeuProTyMetValGluHisThrValGluValLysAsn 626
Db 19058 CAACATTTAAACACTGCTCAAAAGATAAATTTGAACAACAAGTGAACACACACAAAT 19117
QY 627 GlyAsnLeuIleIleProHisLysAspHisTyHisAsnIleLysPheAlaTrpPheAsp 646
Db 19118 -----GTTGAGGTGTAGAT 19132
QY 647 AspHisThrTyLysAlaProAsnGlyTyThrLeuGluAspLeuPheAlaThrIleLys 666
Db 19133 -----ACTGTTAATCA-----AGTGCCACACATTAATGTGTATGGGTACGTTAAGA 19183
QY 667 TyTrpValGlu-----HisProAspGluArgPro 676
Db 19184 AATAGCATACAAGATAACACAGCTACGAAAAATGCGCAAAACTATCTTGATGCTACAGAA 19243
QY 677 HisSerAsnAspGlyTyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSer 696
Db 19244 CGTAAACAAAACAACATAACATGTGTGTGATAGTCTAATGTGTATTAATGAACA 19303
QY 697 GluAspProAsnLysAsnPheLysAla-----Asp 706
Db 19304 AGCAATCCAATATGATGCTGATGCTAATTAACCAATCCCTACACAACTGACATCAACG 19363
QY 707 GluGluProValGluGluThrProAlaGluProGluValProGlnValGluThrGluLys 726
Db 19364 AAAAAATGCTATGATGCTACACATAATTTAACCGACGCAAAACAACAGCAACAATGCC 19423
QY 727 ValGluAlaGlnLeuLys-----GluAlaGluValLeuLeuAlaLysValThr 742
Db 19424 ATCGATGGTGCTACTAATTAATAAGCGCAAAAGATGCGTTAAAGACCAAGTTACA 19483
QY 743 AspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnLeuThr 762
Db 19484 AGTGCGCAACGCTTTCGAATGTAAACA-----AGTATCCAACAACACTGCMAATGAACITTAAT 19540
QY 763 -----LeuGln-----IleMetAspAsnAsnSer----- 770
Db 19541 ACAGCTATGGTCAATTAACACATGTTATGATGATAAATGCAACAAAACAACTCAA 19600
QY 771 -----IleMetAlaGluAla 775
Db 19601 AATATCGTACGCTGAACAAAGTAAGAAAACCTGCTTATGATCAAGCTGTAGTGTGCG 19660
QY 776 GluLysLeuLeuAlaLeuLeuLysGlySerAsnProSerSerValSerLysGluLys 794
Db 19661 AAAGCAATTTTAAATAAACAACAGGTTCAATTCAGATAAAGCAGCGTTGACCGT 19717

RESULT 23

US-09-417-197-108
; Sequence 108, Application US/09417197
; Patent No. 6518021
; GENERAL INFORMATION:
; APPLICANT: Ole THASTRUP, et al.
; TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An Ir
; TITLE OF INVENTION: On A Cellular Response
; FILE REFERENCE: 3759-0110P
; CURRENT APPLICATION NUMBER: US/09/417,197
; CURRENT FILING DATE: 1999-10-07


```

QY 604 sGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHisThrValGluVa 624
Db 1877 -----
QY 624 lLysAsnGlyAsnLeuIlelleProHisLysAspHisTyrHisAsnIleLysPheAlaTr 644
Db 1878 -----CAGGAGAAGGCAGACACGGAAGAGATGATGCCG 1912
QY 644 pPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAlaTh 664
Db 1913 AGGCGCAGATCAT-----GCACGAGCTGGACAACCCCTA----- 1946
QY 664 rIleLysTyrTyrValGluHis-----ProAspGluArgProHisSer-----As 679
Db 1947 -----CATCGTGGGCTCATTTGGCGTCTGCCAGCGAGGCCCTCATGCTGGTCATGG 1999
QY 679 nAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspPr 699
Db 2000 AGATGGCTGGGGGGGGCGCTGCACAAAGTTCTTGTCGGCAAGAGGAGGATCCCTG 2059
QY 699 oAsnLysAsnPhelysAla-----AspGluGluProValGluGl 712
Db 2060 TGACAATGTGGCGAGCTGCTGCACCAAGTGTCCATGGGATGAAGTACCTGGAGGAGA 2119
QY 712 uThr-----ProAlaGluProGluValPro 720
Db 2120 AGAACTTTGTCCACGCTGACCTGGCGGCCCGCAACAGTCTCT 2159

RESULT 24
US-08-961-527-171
; Sequence 171, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 171:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14736 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-171
Alignment Scores: 0.00259 Length: 14736
Pred. No.:

```

```

Score: 148.00 Matches: 158
Percent Similarity: 33.54% Conservative: 109
Best Local Similarity: 19.85% Mismatches: 287
Query Match: 3.55% Indels: 242
DB: 4 Gaps: 38

US-09-765-271-56 (1-796) x US-08-961-527-171 (1-14736)

QY 107 lLysValAspGlyLysTyrTyrVal---TyrLeuLysAspAlaAlaHisAlaAspAsn 125
Db 9349 CTGAAATCGAAGGTATCAATATATTTGTTATATCAA---ACTAAGAAACAGATATAT 9405
QY 126 ValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSerGlnHisArgGluGly 145
Db 9406 ACAGAGCTTTCAAGGACACAGTTGATGGGAATACTCTCTCAAGAGATAGTCAACCAAC 9465
QY 146 GlyThrProArgAsnAsp-----Gly 152
Db 9466 TCTACAAAACATCAGATGTAGTTCACTTCACTGATTTAGATGGAAACCAAGGACAGGG 9525
QY 153 AlaValAlaLeuAlaArgSerGlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsn 172
Db 9526 AAGGTAGTTTA-----CAAGGTCAAGCATCAGGGGATGATGGA----- 9564
QY 173 AlaSerAspIlelleGluAspThrGlyAspAlaTyrIleValProHisGlyAspHisTyr 192
Db 9564 ----- 9564
QY 193 HisTyrIleProLysAsnGluLeuSerAlaSerGluLeuAlaAlaGluAlaPheLeu 212
Db 9565 ---CTTTCAGAAAATCTTCTATAGCAGACAACTATCTCTTAATGATTCATTCGCA 9621
QY 213 SerGlyArgGlyAsnLeuSerAsnSerArgThrTyrArgGlnAsnSerAspAsnThr 232
Db 9622 AGT-----CAAGTTGACGAGATCCGATCACAA 9651
QY 233 SerArgThrAsnTrpValProSerValSerAsnProGlyThrThrAsnThrSer 252
Db 9652 GGAGATCTGTAGTTCGACCAACAGTCCAGAACAGGAAATCTGTCTCTCTACAACG 9711
QY 253 AsnAsnSerAsnThrAsnSerGlnAlaSerGlnSerAsnAspIleAspSerLeuLys 272
Db 9712 GTCAGAGTCCGAGAGAGAGTATTGCGCAGCAAAATGAT-----CGA 9756
QY 273 GlnLeuTyrLysLeuProLeuSerGlnArgHisValGluSerAspGlyLeuValPheAsp 292
Db 9757 CCAGAGTATAAACTTCCATTGGAAACCAAGAGGACGCAAGAACCCGTCATGAGGTGAA 9816
QY 293 -----ProAlaGlnIleThrSerArgThrAlaArgGlyValAla 305
Db 9817 GCCGAGTCCGTGAAGACTTACCAGTCTACACTAAGCCACTAGAAACCAAGGTACACAA 9876
QY 306 ValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMetSerGlnLeuGluGlu 325
Db 9877 GGACCCGGACATGAA-----GGTGAAGCTGCAGTTCGCGAGGAA 9915
QY 326 ArgIleAlaArgIlelleProLeuArg-----TyrArgSer 337
Db 9916 GAACCCAGCTTACACAGAACCGTTAGCAACGAAAGGACGCAAGCCAGGTCTATGAGGC 9975
QY 338 AsnHisTrpValProAspSerArgProGluGlnProSerPro-----GlnProThr 354
Db 9976 AAAGCTACAGTCCGCGAAGAGACTCTTAGAGTACACGGAACCGGTAGCGAAGGACACA 10035
QY 355 ProGluProSerProGlyProGlnProAla-----ProAsnLeuLysIle 369
Db 10036 CAGAACCAGCATCAGGCGGAGCGGAGAGAGAAAGAACTTCGCGCTTTAGAGTC 10095
QY 370 AspSerAsnSerSerLeuValSerGlnLeuVal----- 380
Db 10096 ACTACAGAAATAGAACGGAATCCAGATATTTCTTATACACAGAAATTCAGGAT 10155
QY 381 -----ArgLysValGly----- 384

```

```
Db 10156 CCAACACTTCTGAAAAATCGTGAAGATTGAACGACGAGGCAAGCGGACACGCTACA 10215
QY 385 -----GluGlyTyrValPheGluGluGlyGlySerArg 396
Db 10216 ATTCATATGAGACTACATCGTAATGTTATGTCGTAGAACTATAAGAAAGTGTACGA 10275
QY 397 TyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSer 416
Db 10276 ACTGAAGTAGCT-----CCGCTCAACGAAGTCGTTAAAGTAGGACACACTTGTGAAA 10326
QY 417 LysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAsp 436
Db 10327 GTTAAACCTCAGTAGAATAATCAACACTTAAACAAAGTTTGAAC----- 10371
QY 437 GlnGluPheTyrAspLysAlaTyrAsnLeuLeu-----ThrGluAlaHisLysAlaLeu 454
Db 10372 AAAAAATCTATAACTGTAAGTTATATACTTAATAGACACTTACCTCAGCATATGTTCT--- 10428
QY 455 Phe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuGluArgLeu 474
Db 10429 -----GCAAAACGCAAGTCTTTCCATGGA---GACAACTAGTTAAAGAGGTG 10473
QY 475 AsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaPro 494
Db 10474 GATATAGAAATCTCTGCCAAGACGCAAGTAATATCAGGT---TTAGATTACTACACACG 10530
QY 495 Ile-----ThrHis---ProGluArgLeuGlyLysProAsnSerGlnIleGluTyr 510
Db 10531 TATACAGTTTAAACACACACCTAACTATATATTTGGGTGAAAATAATCAG-----GAAAAT 10584
QY 511 ThrGluAspGluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyr 530
Db 10585 ACTGAACATCAACTCAAGATTCCAAATTA-----GAGTAAGAAATAAGAGATTAAA 10638
QY 531 IlePheAspGluHisAspIleLeuSerAspGluGlyAspAlaTyrValThrProHisMet 550
Db 10639 GATATTGATTCAGTAGAATTATACGTAAGAAATGATCGTTAT----- 10683
QY 551 GlyHisSerHisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaGln 570
Db 10684 -----CGTAGATATTTA-----AGTCTAAGTGAAGCGCGACTGATACGGCTAAA 10728
QY 571 AlaTyrThrLysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsn 590
Db 10729 TACTTTGTAAAGTGAAA----- 10746
QY 591 ProThrGlyAspSerAlaAlaIleTyrAsnArgValLysGlyGluLysArgIlePro 610
Db 10747 -----TCAGATCGCTTCAAGAAATGTACCTACCTGTAAATCT----- 10785
QY 611 LeuValArgLeuProTyrMetValGluHisThr-----ValGluValLys 625
Db 10786 -----ATTACAGAAATACCGATGGAACGATATAAGTGAACGGTAGCC 10827
QY 626 AsnGlyAsnLeuIleLeuProHisLysAspHisTyrHisAsnIleLysPheAlaTyrPhe 645
Db 10828 GTTGATCAACTTCTCGAAGAGGTACAGCGTTACAAA----- 10866
QY 646 AspAspHisThrTyr-----LysAlaProAsnGlyTyrThr--- 657
Db 10867 GATGATTACATATTACTGTAGCTAAATCTAAAGCAGACGCAACAGGAGTTTACACATCC 10926
QY 658 LeuGluAspLeuPheAlaThrIleLys-----TyrTyrValGluHisPro 672
Db 10927 TTTTAAACAGCTGTGAACAGCAGTCAAGCAATCTGTCTGGTGTCTATACATTGGCTTCA 10986
QY 673 AspGluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLys 692
Db 10987 GATATG-----ACCGCATGAGGTGACCTTAGGCGAT 11019
QY 693 LysAspHisSer-----GluAspProAsn 700
```

```
Db 11020 AAGCAGACAAGTTATCTCACAGGTGCATTTTACAGGAGCTTGATCGGTTCTGATGGAAACA 11079
QY 701 LysAsnPheLysAlaAspGlu-----GluProValGluGluThrProAlaGluProGlu 718
Db 11080 AAATCGTATGCACTTATGATTGGAAGAACCATTAATTGATACATTAAATGCTGCTACA 11139
QY 719 ValProGlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeu 738
Db 11140 GTTAGAGATTGGATTAATAAAGCTGTTCTCTGATAGTAAAGAAAATGTCGACGCGCTG 11199
QY 739 AlaLysValThrAspSerSerLeu-----LysAlaAsnAla 750
Db 11200 CGCAAGGCGAGCAATAGCGCAATTAATAATGTTTCAGTAGAAGGAAAATCTCAGGT 11259
QY 751 ThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnSer 770
Db 11260 CGCAAAATCTGTTCGGGATTAGTAGCGAGCGCAACAATACAGTATAGATAAAGACGCTCG 11319
QY 771 IleMetAlaGluAlaGluLysLeuLeuAlaLeuLeuLysGlySerAsn 786
Db 11320 TTTACAGGG-----AACTTATCGCAATCACCAGGACAGTAAT 11358
RESULT 25
US-09-817-514A-5
; Sequence 5, Application US/09817514A
; Patent No. 6639129
; GENERAL INFORMATION:
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; CURRENT FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 2745
; TYPE: DNA
; ORGANISM: Photorhabdus luminescens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2745)
US-09-817-514A-5
Alignment Scores:
Pred. No.: 0.000191 Length: 2745
Score: 147.50 Matches: 178
Percent Similarity: 32.99% Conservative: 112
Best Local Similarity: 20.25% Mismatches: 380
Query Match: 3.54% Indels: 210
DB: 4 Gaps: 35
US-09-765-271-56 (1-796) x US-09-817-514A-5 (1-2745)
QY 13 LysGluAsnAsnArgValSerTyrIleAsp-GlyLysGlnAlaThrGlnLysThrGluAs 32
Db 408 AGAGAAAACGACGGAACGCTTATCTGGCCGCGCAATACGCCGCAAGAAAAGATTACAA 467
QY 32 nLeuThrProAspGluValSerLysArgGlu----- 42
Db 468 CCTCGCGGTCACTGTGTCGCCCATTCAGTACCGCGGACCTTACTCACTCAATAGCCT 527
QY 43 -----GlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVa 59
Db 528 TTCTCTGGCTGGCTGCTGCTATCAATCTCAACACTCTCAACCTCCTACCATCAACAGGATGC 587
QY 59 lThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleI 79
Db 588 CGACTGCAGGTTGAAGACACGAGCCCTCTGGCAACAAAACACTGATGATGTCTATAT 647
```

QY 79 eSerGluGlu-----LeuLeuMetLysAspPro-----As 89
 Db |:::|
 648 CACCCAAAGTACACTGATGCCACGGGCTTACTGACCCAGCATGCCAAAGGCAA 707
 QY 89 nTyrLysLeuLysAspGluAspIleValAsnGluValLysGlyTyrValIleLysVa 109
 Db |:::|
 708 CATTGAGCGGCTGCGCTATGATGCGCGGCGAGCTAAAGAGGAGTTGTTAACTCAA 767
 QY 109 lAspGlyLysTyrTyrValTyrLeuLysAspAlaIleHis-----AlaAspAsnValar 127
 Db |:::|
 768 AGTGCAGCGGACAGAGGTGATTATCAATTCGCTAACTCTCCGCGCGGCGCAAAAAAT 827
 QY 127 gThrLysGluGluIleAsnArgGlnLysGlnLysGlnHisArgGluGlyGlyTh 147
 Db |:::|
 828 AGTGAAGAGCAGCAGTACGGGATGTCAGTGAATACAGCTACGACCGGAA-----AC 881
 QY 147 rProArgAsnAspGlyAlaValAlaLeuAlaArgSerGlnGlyArgTyrThrAspAs 167
 Db |:::|
 882 CCAACGGCTTATCGCATACCATCGCGCTCCATCAGACGCCAAGGTTGCAAGACCT 941
 QY 167 pGlyTyrIlePheAsn---AlaSerAspIleIleGluAspThrGlyAspAlaTyrIleVa 186
 Db |:::|
 942 ACGTATCAATATACCCAGTAGGCAATGTCATTAATATCCGTATATCGTGAAGGCCAC 1001
 QY 186 lProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeuSerAlaSerGluLeu-- 205
 Db |:::|
 1002 TCGCTTTTGGCGCAATCAGAAAGTAGCCCGGAGATAGCTATACCTACGATCCCTGTA 1061
 QY 206 -----AlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSerAr 222
 Db |:::|
 1062 TCAGCTTATCAGCCGCCACCGCGCGGAAATCGCCAAATATCGTCAGCAAAACCACT 1121
 QY 222 gThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSerVal-- 241
 Db |:::|
 1122 TCCCTCCCTCGGTACTCTCTGACAAATATCTACCTAACTATATCTCGCAGCTACAG 1181
 QY 242 -----SerAsnProGlyThrThrAsn-- 248
 Db |:::|
 1182 CTATGATCACAGTGGTAATCTGACGCAAAATTCGCGCACAGCTCGCCAGCTACCCAGAACAA 1241
 QY 249 -----ThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAl 261
 Db |:::|
 1242 CTACACCGTGGCTATACCCCTCTCAAAACCGCAGCAATCGGGGTCTCTCAGTACGCTAAC 1301
 QY 261 aSerGlnSerAsnAspIleAspSerLeuLeu-----LysGlnLeuTyrLysLe 277
 Db |:::|
 1302 CACGGATCCAAATCAAGTGGATAGTTGTTGATGCGCGGTGGTCCACCAACCCAGTTTATT 1361
 QY 277 uProLeuSerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleTh 297
 Db |:::|
 1362 ACCC-----GGACAGACACTTATCTGGACACCACGAGGAGGTT 1400
 QY 297 rSerArgThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTy 317
 Db |:::|
 1401 AAGCGGTTAAATATGGC-----CCGGAAATGAGTGTAC----- 1437
 QY 317 rSerGlnMetSerGluLeuGluArgIleAlaArgIleIleProLeuArgTyrArgSe 337
 Db |:::|
 1438 -----CGTACGACAG 1448
 QY 337 rAsnHisTrpValProAspSerArgProGluGlnProSerProGlnProThrProGluPr 357
 Db |:::|
 1449 CAACGGCATGAGACAACTGAAGTGAGTGAACAGCCACCCAGCAATACTACGAGCAACA 1508
 QY 357 oSerProGlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSerSerLeuVa 376
 Db |:::|
 1509 ACGGGTAATCTATTTCGCGGAGTGGAGCTTACGCAACCCAGACGCAACGCCAACAAAC 1568
 QY 376 lSerGlnLeu---ValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSe 395
 Db |:::|
 1569 GGNAGAGTTACAGTTATCACACTCGGTGAAGCC-----GGTCCGCGCACAGGT 1616

QY 395 rArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysSe 415
 Db |:::|
 1617 ACGGGTGTTCACCTGGGAGCGGTAAAGCAGAGATGTCAACAATAATCACTACGTTA 1676
 QY 415 uSerLysGlnGluSerVal---SerHisThrLeuThrAlaLysLysGluAsnValAlaPr 434
 Db |:::|
 1677 CAGCTACCATATCTGATCGGCTCCAGCAGCTTGAAGTGGCAACAGGACAAATATT 1736
 QY 434 oArgAspGlnGluPheTyrAsp-----LysAlaTyrAsnLeuLeuTh 448
 Db |:::|
 1737 CAGCGAGGAAGAGTATTATCCATTGGCGGACACGCGTGTGGCAGCAACAGCAACAA 1796
 QY 448 rGluAlaHisLysAlaLeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAs 468
 Db |:::|
 1797 AGAAGCCAGCTATATAAACGATTCGCTATTCGCGCAAGACGAGATGCCCGGTTGTA 1856
 QY 468 pLysLeuLeuGluArgLeuAsnAspGlu-----SerThrAsnLysGlu 482
 Db |:::|
 1857 TTATTACGGTTATCGTTATTACCAACCGTGGCGGCGAGATGGTTAAGCGGAGCCCGC 1916
 QY 482 uLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArgLeuGl 502
 Db |:::|
 1917 AGGAACCATTCGATGGCTGAATCTATACCGAATGTAAGAAATAATCTCTGTGAGTTTA-- 1974
 QY 502 yLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeuAlaAs 522
 Db |:::|
 1975 -----CAAGATGAATAATGGATTAGCGCCAGAAAAAGG 2006
 QY 522 pLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAspGluGl 542
 Db |:::|
 2007 GAATATACCAAGAGGTAAATTTC---TTGATGAATTAAATCAAAATTCGAATGGCAGCAA 2063
 QY 542 yAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSerLeuSe 562
 Db |:::|
 2064 AAGTTCACATGTTGTC-----AAATGGAACGAGAAAGAGAC-- 2100
 QY 562 rAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIle---LeuProPr 581
 Db |:::|
 2101 -----AGTTATACAAAAATAAATCAATGAAAGTGGTTCG 2135
 QY 581 oSerProAspAlaAspValLysAlaAsnProThrGly----- 593
 Db |:::|
 2136 TGTCGGTGAATCCGAT-----CCGTGGGTATTTCCTAAGCCACGAAGATT 2183
 QY 594 -----AspSerAlaAlaIleTyrAsnArg-Val----- 603
 Db |:::|
 2184 ACTAAAGGTATAGAAAAAGTCAATATATAGCCGACTTGAAGAAACAGCTCCCT 2243
 QY 604 -----LysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHisThrV 622
 Db |:::|
 2244 TTCAGAAAAATCAAAACCGAATCTTTCTTTAGGATCTGAAATATCCGGTTATATGGAAG 2303
 QY 622 alGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIleLysP 642
 Db |:::|
 2304 AA-----CCATACAGATACGATATCAGATATGCC-- 2334
 QY 642 heLaTrpPheAspAspHisThrTyrLysAla-----ProAsnGlyTyrThrLeuGluA 660
 Db |:::|
 2335 -----GAAGAGCATATAAGAAAGTAAATCACCTGTATTTTATTTCAGAAACCG 2383
 QY 660 spLeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnA 680
 Db |:::|
 2384 ATTCTTTCGGTTAATG-----GATAAAGTGAATAAATG 2419
 QY 680 spGlyTrpGlyAsnAlaSerGluHis---ValLeuGlyLysLysAspHisSerGluAspP 699
 Db |:::|
 2420 ATTATTCCGGTGAAGAAAAATTTATCGCGCAATGAGGTTTAAAGTTTATCATGATTAA 2479
 QY 699 roAsnLysAsnPheLysAlaAspGluProValGluGluThrProAlaGluProGluV 719
 Db |:::|
 2480 AAAATAAAACAA-----TCAGAATTACATGTCAACTATGCATTTGCGCCATCCC--- 2526
 QY 719 alProGlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuA 739

Db	2527	-----TATACCGCAATTGAGTAATGAACAAAGACGCTGT	2560
Qy	739	laLysValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgA	759
Db	2561	TGCAGAAACACAGAACCCGCTATTGCAATAGATAGAGATATAATTTCAAAGGTGTTGGCA	2620
Qy	759	snAsnLeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuL	779
Db	2621	AATTCCTGACAATGAAGCAATTAAAAATCA-----	2652
Qy	779	euAlaLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn	796
Db	2653	-----TTGAAGGACATAAAATTAAATAGATATCAACAGAGCTATTAAAT	2697

Search completed: October 1, 2004, 11:34:14
Job time : 363 secs


```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:      October 1, 2004, 10:19:33 ; Search time 928 Seconds
           (without alignments)
           4348.465 Million cell updates/sec

Title:       US-09-765-271-56
Perfect score: 4165
Sequence:    1 SYEGLYQARTVKENRVS.....KLLALLKGNPSSVSKEKIN 796

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched:    3340653 seqs, 2534783454 residues

Total number of hits satisfying chosen parameters:      6681306

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 90 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool.h/US09765271/runat_30092004_113756_12374/app_query.fasta_1.967
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=90 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=25 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09765271@cgn.1.1.593 @runat_30092004_113756_12374
-NCPU=6 -ICPU=3 -NO.MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LOGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result      *      Query
No.         Score Match Length DB ID      Description

```

1	4163	100.0	2389	9	US-09-765-272-55	Sequence 55, Appl
2	4163	100.0	2451	13	US-10-412-862-9	Sequence 9, Appl
3	4163	100.0	2451	16	US-10-412-850-9	Sequence 9, Appl
4	4163	100.0	2451	16	US-10-387-783-9	Sequence 9, Appl
5	4163	100.0	8195	13	US-10-158-844-94	Sequence 94, Appl
6	3218	77.3	2523	10	US-09-884-465A-3	Sequence 3, Appl
7	3218	77.3	2647	10	US-09-884-465A-4	Sequence 4, Appl
8	2795	67.1	2481	10	US-09-769-787-206	Sequence 206, App
9	2786.5	66.9	2531	13	US-10-412-862-11	Sequence 11, Appl
10	2786.5	66.9	2531	16	US-10-412-850-11	Sequence 11, Appl
11	2786.5	66.9	2531	16	US-10-387-783-11	Sequence 11, Appl
12	2772	66.6	2531	16	US-10-412-862-5	Sequence 5, Appl
13	2772	66.6	2531	16	US-10-412-850-5	Sequence 5, Appl
14	2772	66.6	2531	16	US-10-387-783-5	Sequence 5, Appl
15	2750.5	66.0	2639	10	US-09-884-465A-5	Sequence 5, Appl
16	2649.5	63.6	2290	9	US-09-765-272-65	Sequence 65, Appl
17	1774	42.6	2359	13	US-10-158-844-243	Sequence 243, App
18	1247	29.9	3120	10	US-09-884-465A-1	Sequence 1, Appl
19	1247	29.9	5048	10	US-09-884-465A-2	Sequence 2, Appl
20	1222.5	29.4	6867	13	US-10-158-844-192	Sequence 192, App
21	1203.5	28.9	1455	10	US-09-769-787-246	Sequence 246, App
22	1203.5	28.9	1455	10	US-09-769-744A-23	Sequence 23, Appl
23	1203.5	28.9	1455	13	US-10-412-862-7	Sequence 7, Appl
24	1203.5	28.9	1455	16	US-10-412-850-7	Sequence 7, Appl
25	1203.5	28.9	1455	16	US-10-387-783-7	Sequence 7, Appl
26	1187.5	28.5	1342	9	US-09-765-272-181	Sequence 181, App
27	988	23.7	973	13	US-10-158-844-355	Sequence 355, App
28	966.5	23.2	2528	10	US-09-884-465A-9	Sequence 9, Appl
29	946	22.7	5215	10	US-09-252-088-13	Sequence 13, Appl
30	946	22.7	5215	16	US-10-340-792-13	Sequence 13, Appl
31	929	22.3	2469	10	US-09-769-736-17	Sequence 17, Appl
32	841	20.2	1146	10	US-09-769-736-23	Sequence 23, Appl
33	770.5	18.5	1684	13	US-10-158-844-258	Sequence 258, App
34	650	15.6	2721	16	US-10-324-143-6	Sequence 6, Appl
35	644	15.5	816	16	US-10-324-143-4	Sequence 4, Appl
36	644	15.5	816	16	US-10-324-143-5	Sequence 5, Appl
37	643	15.4	819	10	US-09-884-465A-257	Sequence 257, App
38	643	15.4	819	16	US-10-324-143-10	Sequence 10, Appl
39	512.5	12.3	1455	10	US-09-769-736-71	Sequence 71, Appl
40	304.5	7.3	841	9	US-09-452-599-34	Sequence 34, Appl
41	304.5	7.3	841	15	US-10-121-120-34	Sequence 34, Appl
42	172	4.1	7446	13	US-10-282-122A-7578	Sequence 7578, Ap
43	170	4.1	7434	9	US-09-815-242-4761	Sequence 4761, Ap
44	170	4.1	7437	9	US-09-815-242-8869	Sequence 8869, Ap
45	170	4.1	8155	8	US-08-781-986A-63	Sequence 63, Appl
46	170	4.1	8155	13	US-10-329-624-63	Sequence 63, Appl
47	169	4.1	2730	15	US-10-172-502-5	Sequence 53, Appl
48	169	4.1	12173	8	US-08-781-986A-310	Sequence 310, App
49	169	4.1	12173	13	US-10-329-624-310	Sequence 310, App
50	168.5	4.0	3132	9	US-09-946-805-3	Sequence 3, Appl
51	164	3.9	2676	13	US-10-282-122A-7677	Sequence 7677, Ap
52	163	3.9	11076	13	US-10-282-122A-35051	Sequence 35051, A
53	159.5	3.8	14172	16	US-10-359-012-7	Sequence 7, Appl
54	159	3.8	2613	9	US-09-815-242-4438	Sequence 4438, Ap
55	159	3.8	2613	9	US-09-815-242-8539	Sequence 8539, Ap
56	159	3.8	19922	13	US-10-282-122A-34396	Sequence 34396, A
57	158	3.8	4359	10	US-09-873-367C-94	Sequence 94, Appl
58	158	3.8	4359	15	US-10-153-668-103	Sequence 103, App
59	157	3.8	5419	17	US-10-479-546-3	Sequence 3, Appl
60	157	3.8	7088	17	US-10-451-467A-383	Sequence 383, App
61	156	3.7	1766	10	US-09-934-455-163	Sequence 163, App
62	156	3.7	1766	13	US-10-412-699B-697	Sequence 697, App
63	156	3.7	1766	13	US-10-225-066A-599	Sequence 599, App
64	156	3.7	1766	16	US-10-302-267-35	Sequence 35, Appl
65	156	3.7	1766	16	US-10-374-780A-2519	Sequence 2519, Ap
66	155.5	3.7	3369	13	US-10-282-122A-34397	Sequence 34397, A
67	154	3.7	5085	15	US-10-198-846-9854	Sequence 9854, Ap
68	153.5	3.7	5967	15	US-10-171-581-158	Sequence 158, App
69	152.5	3.7	7200	13	US-10-282-122A-33992	Sequence 33992, A
70	152.5	3.7	31096	8	US-08-781-986A-59	Sequence 59, Appl
71	152.5	3.7	31096	13	US-10-329-624-59	Sequence 59, Appl
72	151.5	3.6	6258	9	US-09-815-242-4584	Sequence 4584, Ap

Sequence 8512, Ap
Sequence 1, Appli
Sequence 108, App
Sequence 1, Appli
Sequence 7653, Ap
Sequence 7627, Ap
Sequence 171, App
Sequence 5, Appli
Sequence 26, Appl
Sequence 5, Appli
Sequence 32569, A
Sequence 45972, A
Sequence 4580, Ap
Sequence 8291, Ap
Sequence 43, Appl
Sequence 33842, A
Sequence 33993, A
Sequence 38279, A

ALIGNMENTS

RESULT 1
US-09-765-272-55
; Sequence 55, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272

FILING DATE: 22-Jan-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 55:

SEQUENCE CHARACTERISTICS:

LENGTH: 2389 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 55:

US-09-765-272-55

Alignment Scores:
Pred. No.: 0
Score: 4163.00
Best Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 99.95%
DB: 9

US-09-765-271-56 (1-796) x US-09-765-272-55 (1-2389)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVallysGluAsnAsnArgValSerTyr 20
Db 2 TCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAATAATACGTTGTTCTTAT 61
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 62 ATAGATGAAACCAAGCGACGCAAAACGAGAAATTTGACTCTCTGATGAGGTTAGCAAG 121
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db 122 CGTGAAGAAATCAATGCTGAGCAATCTCATCAAGATACAGACCAAGGCTATGCTACT 181
QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleSer 80
Db 182 TCACATGCGGACCACTATCATTTACATGTTAAAGGTTCTTATGACGCTATCATCAGT 241
QY 81 GluGluLeuLeuMetLysAspProAsnTyrIlysLeuLysAspGluAspIleValAsnGlu 100
Db 242 GAAGAATTAATCATGAAAGATCCAACTATTAAGCTTAAAGATGAGGATATTGTTAATGAG 301
QY 101 VallysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrValTyrLeuLysAspAla 120
Db 302 GTCAAGGGTGGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGCT 361
QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
Db 362 GCCACGCGGGATAACGTCGGTACAAAGAGGAAATCAATCGACAAACCAAGAGCATAGT 421
QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaIleAlaIleAlaSerGln 160
Db 422 CAACATCGTGAAGTGGAACTCCAAAGAACGATGCTGCTGTTGCTTGGCAGCTTCGCAA 481
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db 482 GGACGCTATPACTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGATACT 541
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db 542 GGTGATGCTTATATGCTTCTCATGAGATCATTAACATTACATTCTAAGAATGAGTTA 601
QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyValAsnLeuSerAsn 220
Db 602 TCAGCTAGCGAGTTGGCTGCTCAGAAAGCTTCTCTATCTGGTGGAGAAATCTGTCAAT 661
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
Db 662 TCAAGAACCTATCGCGCAAAATACGATACACTTCAAGAACAACTGGGTACCTTCT 721
QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db 722 GTAAGCAATCCAGGAACCTACAAATATAACACAAAGCAACACACAGCAACACTACAGTCRA 781
QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGluLeuTyrLysLeuProLeuSer 280
Db 782 GCAAGTCAAGTAAATGACATTTGATGCTCTTGAACACAGCTCTACAAACTGCTTTGAGT 841
QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db 842 CAACGACATGTAGAAATCTGATGCGCTTGTCTTTGATCCAGCACAAATACCAAGTCGAACA 901
QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db 902 GCTAGAGGTTGTCAGTGCCACACGAGATCATTAACCATCTTCACTTCTTCTTCTCAATG 961
QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyrp 340
Db 962 TCTGAATGGAAGAACGAATCGCTCGTATTATTTCCCTTCTGTTATCGTTCAACCATGG 1021
QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db 1022 GTACCATTTCAAGGCCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGGC 1081

```

QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
Db 1082 CGCAACCTCGCACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTAGTCAGCTGGTA 1141
QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
Db 1142 CGAAAGTTGGGAAGGATATGTTTCGAAGAAAGGGCATCTCTCGTTATGCTTTGGC 1201
QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
Db 1202 AAAGATTTACCACTCTGAACCTGTTAAATCTTTGAAGCAAGTTATCAAAAACAAGAGAT 1261
QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
Db 1262 GTTTCACACACTTTAACTGCTAAAGAAAGAAATGTTGCTCTCTGTCGACCAAGATTTAT 1321
QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db 1322 GATAAAGCATATAATCTGTTAACTGAGGCTCATAAGGCTTTGTTGNAATAAAGGTCGT 1381
QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
Db 1382 AATCTGATTTCCAAAGCTTTAGACAAATATTAGAACGCTTGAATGATGAATCGACTAAT 1441
QY 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db 1442 AAAGAAAAATTTGGTAGATGATTTATTGGCAATCTCTAGCACCAATYACCCATCCAGAGCA 1501
QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
Db 1502 CTTGGCAAAACCAAAATCTCAAAATTTAGTATATCTGAAGACGAATTCGTATTGCTCAATTA 1561
QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
Db 1562 GCTGATAGTATACACGTCAGATGGTTTACATTTTGTGATGAACATGATATATCAATCAGTAT 1621
QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisThrIleGlyLysAspSer 560
Db 1622 GAAGGAGATGCATATGTAAGCCCTCATATGGCCATAGTCACTGGATTTGNAAGATAGC 1681
QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
Db 1682 CTTTCTCATAGGAAAAAGTTGACGCTCAAGCCTATATCAAGAAAAAAGGTATCCTACCT 1741
QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaAlaIleTyr 600
Db 1742 CCATCTCCAGACGAGATGTTAAGCAAAATCCAACTGGAGATAGTCAGCAGCTATTATTC 1801
QY 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
Db 1802 AATCGTGTGAAGGGGAAAAACCAATTCCTCGTTCGACTTCATATATGTTGAGCAT 1861
QY 621 ThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIle 640
Db 1862 ACAGTTGAGGTTTAAAAACGGTAAATTTGATTAATTTCTCATAGGATCATTAACCAATAAT 1921
QY 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660
Db 1922 AAATTTGCTTTGTTGATGATCACAATCAAAAGCTCCAAATGGCTATATACCTTTGGAAGAT 1981
QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
Db 1982 TTGTTTGGCAGCATTAAGTACTAGTAGAACACCCCTGACCAAGCTCCACCATTTATGAT 2041
QY 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
Db 2042 GGATGGGGCAATCCGATGAGCATGTGTTAGGCAAGAAAGACACACAGTGAAGATCCAAAT 2101
QY 701 LysAsnPhelysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720
Db 2102 AAGAACTTCAAGCGGATGAAGCCAGTAGAGGAAACACCTCTGAGCCAGAGTCCCT 2161

```

```

QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
Db 2162 CAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAGAGAGCAGAAAGTTTGTCTTGGAAA 2221
QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
Db 2222 GTAACGGGATTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAATAAT 2281
QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuAla 780
Db 2282 TTGACTCTTCAATTTATGATAACAATAGTATCATATGGCAGAGCAGAAAAATTTACTTGGC 2341
QY 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796
Db 2342 TTGTTAAAGGAAGTAATCTCTTCATCTCTAAGTAGGAAAAATAAAC 2389

RESULT 2
US-10-412-862-9
; Sequence 9, Application US/10412862
; Publication No. US20040052781A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-685
; CURRENT APPLICATION NUMBER: US/10/412,862
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 9
; LENGTH: 2451
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(2451)
; OTHER INFORMATION: n = a, c, t or g
US-10-412-862-9

Alignment Scores:
Pred. No.: 0 Length: 2451
Score: 4163.00 Matches: 796
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.95% Indels: 0
DB: 13 Gaps: 0

US-09-765-271-56 (1-796) x US-10-412-862-9 (1-2451)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db 61 TCTTACGAGTTGGAGCTGTATCAAGCTTAGAAGCAATTAATCGTGTTCCTAT 120
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 121 ATAGATGGAAACACACGACGCAAAACCGAATTTGACTCCTCTAGGTTAGCAAG 180
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db 181 CGTGAAGGAATCAATCTCTGAGCAATCGTCATCAAGATAACAGCCAGCTATGTCAC 240
QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db 241 TCACATGGCGACCACTATCATTTACATGTTAGAGTTCTTCTTATCAGCTATCATCAGT 300
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100

```

Db 301 GAAGAAATTACTCATGAAGAAGTCCAAACTATTAAGCTAAAGATGAGGATATTGTTAATCAG 360
 Qy 101 VallysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrValTyrLeuLysAspAla 120
 Db 361 GTCAGGGTGGATATGTTATCAAGGTAGATGGAATACTAATGTTTACTTAAAGATGCT 420
 Qy 121 AlahiSaLaAspAsnValArgThrLysGluGluLysLeuLysGlnLysGlnLysSer 140
 Db 421 GCCACGCGGATAAGCTCGTACAAAGAGGAAATCAATCCAGCAAAAACAAGAGATAGT 480
 Qy 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
 Db 481 CAACATCGTAGGGTGGAACTCCAGAAACGATGGTGTGTGTGCTTGGCAGTTGCGAA 540
 Qy 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
 Db 541 GGACCTATACATACAGATGATGTTATATCTTTAATGCTTCTGTATCATAGAGATACT 600
 Qy 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
 Db 601 GGTGATGCTTATATCGTTCCTCATGAGATCAATTACCATTAATCCCTAAGAAATGAGTTA 660
 Qy 201 SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
 Db 661 TCAGCTAGCGAGTTGGCTGCTGCAGAGCTTCTCTATCTGTGTCGAGGAAATCTGTCAAA 720
 Qy 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
 Db 721 TCAAGAACCTATCGCCGCAAAATAGCGATAACACTTCAAGAACAACTGGGTACTCTCT 780
 Qy 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
 Db 781 GTAAGCAATCCAGGAACTCAAAATATAACACAGCAACACAGCAACACTAACAGTCAA 840
 Qy 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
 Db 841 GCAAGTCAAAAGTAATGACATTGATGCTCTTGAACAGCTCTACAACTGCTTTGAGT 900
 Qy 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
 Db 901 CAACGACATGTAAATCTGATGGCTTGTCTTTGATCCAGCAAAATCAAAAGTCGAAACA 960
 Qy 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
 Db 961 GCTAGAGGTGTTCAGTGGCCACAGGAGATCATTCACCTTATCCCTTACTCTCAATG 1020
 Qy 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340
 Db 1021 TCTGAATTGGAAGACGAATCGCTCGTATATTTCCTTCTGTTATCGTTCAAAACCATGG 1080
 Qy 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
 Db 1081 GTACCAAGATTCAAGGCGCAGAACCAACCAAGTCCAAACCGACTCCGGAACCTAGTCAGGC 1140
 Qy 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
 Db 1141 CCGCAACCTGCACCAAACTTTAAATAGACTCAAAATCTCTTTGGTGTAGTCACTGTA 1200
 Qy 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
 Db 1201 CGAAAGTTGGGGAAGGATATGATTTCGAAGAAAAGGGCATCTCTCGTTATGCTTTGCG 1260
 Qy 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLysLeuSerLysGlnGluSer 420
 Db 1261 AAAGATTATACCATCTGAACCTGTTAAATAATCTTGAAGCAAGTATCAAAACCAAGAGAT 1320
 Qy 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
 Db 1321 GTTTCACACACTTTAATGCTTAAAGAAAAGTGTCTCTCTCGTACCAAGAAATTTAT 1380
 Qy 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
 Db 1381 GATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATAAGGGTCT 1440

Qy 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
 Db 1441 AATTCTGATTTCCAGGCTTAGACAAATATTAGAACGCTTGAATGATGAATCAGCTAAT 1500
 Qy 481 LysGluLysLeuValAspAspLeuAlaPheLeuAlaProIleThrHisProGluArg 500
 Db 1501 AAAGAAAATTTGGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCTCCAGAGCGA 1560
 Qy 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
 Db 1561 CTTGGCAAAACCAAAATCTCAAAATTGAGTATACTGAAGACGAAGTTCTGTTTGTCTCAATTA 1620
 Qy 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
 Db 1621 GCTGATAGTATATCAACGTCAGATGGTTACATTTTGTATGAACATGATATATCACTGAT 1680
 Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560
 Db 1681 GAAGGAGATGCATATGTAAACGCTCATATGGGCCCATAGTCACTGGATTGGAAAAGATAGC 1740
 Qy 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
 Db 1741 CTTTCTGATAGGAAAAGTTGCGACTCAAGCCTATATCTAAGAAAAGGTATCTCTACT 1800
 Qy 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600
 Db 1801 CCATCTCCAGAGCGAGATGTTAAAGCAAAATCCACTGGAGATAGTGCAGCAGCTATTTAC 1860
 Qy 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
 Db 1861 AATCGTGTAAAGGGGAAAACGAAATCCACTCGTTCGACTTCCATATATGTTGAGCAT 1920
 Qy 621 ThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIle 640
 Db 1921 ACAGTTGAGTTAAAAACGTTAATTTGATTATCTCTCATAGGAATCATCCATAATATT 1980
 Qy 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660
 Db 1981 AATTTGCTTGGTTGATGATCACACATCAAAAGCTCCAAATGGCTATACCTTGAAGAT 2040
 Qy 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
 Db 2041 TTGTTTGGCAGCATTAAGTACTACGTAGAACACCCCTGACGAACGTCACATCTTAATGAT 2100
 Qy 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
 Db 2101 GGATGGGGCAATGCCAGTGCAGCATGTGTTAGGCAAGAAAGACACACAGTGAAGATCCAAAT 2160
 Qy 701 LysAsnPheLysAlaAspGluProValGluProValGluThrProAlaGluProGluValPro 720
 Db 2161 AAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGAAAACACCTCTGAGCCAGAAAGTCCCT 2220
 Qy 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
 Db 2221 CAAAGTAGAGACTGAAAAGTAGAAGCCCACTCAAGAGAGCAGAAAGTTTTGCTTGCAGA 2280
 Qy 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
 Db 2281 GTHAACGATTCTAGTCTGAAGCCATGCAACAGAACTCTAGCTGGTTTACGAAATAAT 2340
 Qy 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780
 Db 2341 TTGACTCTTCAAATTTATGGATAACAATAGTATCATGCGAAGACAGAAAATTTACTTGGC 2400
 Qy 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796
 Db 2401 TTGTTAAAGGAAGTAACTCTTCTATCTGTAAGTAAGGAAAATAATTAAC 2448

RESULT 3
 US-10-412-850-9
 ; Sequence 9, Application US/10412850
 ; Publication No. US2004000186A1

```

; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-686
; CURRENT APPLICATION NUMBER: US/10/412,850
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 9
; LENGTH: 2451
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(2451)
; OTHER INFORMATION: n = a, c, t or g
US-10-412-850-9

Alignment Scores:
Pred. No.: 0 Length: 2451
Score: 4163.00 Matches: 796
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.95% Indels: 0
DB: 16 Gaps: 0

US-09-765-271-56 (1-796) x US-10-412-850-9 (1-2451)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db 61 TCTTACGAGTTGGACTGTATCAAGCTAGAACGGTTAAGGAAATATTCGTTTCCTAT 120
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 121 ATAGATGGAAACAAAGCGACGCAAAACCGAGAAATTTGACTCTCGTAGAGTTAGCAAG 180
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db 181 CGTGAAGGAATCAATGCTGACCAATCGTCAATCAAGTAATACAGACCAAGGCTATGTCAT 240
QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleSer 80
Db 241 TCACATGGCGACCACTATCATATTATCAATCGTTAAGTTCTTATGACGCTATCATCAGT 300
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db 301 GAAGATTACTCATGAAGATCCAACTATAAGCTAAAGATGAGGATATGTTAATGAG 360
QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db 361 GTCAAGGTTGGATATGTTATCAAGTAGATGGAAATACTATGTTTACCTTAGGATGCT 420
QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnHisSer 140
Db 421 GCCCAGCGGATACGTCCTGACAAAGAGCAAAATCAATCAACAAACAAAGACATAGT 480
QY 141 GlnHisArgGluGlyThrProAspAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
Db 481 CAACATCGTGAAGTGAATCCAAAGAAACGATGGTGGTCTGCTGGCACGTTGCGAA 540
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsnThr 180
Db 541 GGACGCTACTACAGATGATGTTATCTTTAATGCTTCTGATATCATAGAGGATACT 600
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200

```

```

Db 601 GGTGATGCTTATATCGTTCTTCATCGAGATCATTTACCATTACATTCCTTAAGAATGAGTTA 660
QY 201 SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db 661 TCAGCTAGCGAGTTGGCTGCTGCAGAACCTTCTTATCTGCTCGAGGAATCTGTCAAT 720
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
Db 721 TCAGAAGCTTATCGCCGCAAAATAGCGATAACACTTCAAGAACAACTGGGTGCTTCT 780
QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db 781 GTAAGCAATCCAGGACTTCAAACTACTAACACCAACCAACAGCAACACTAACAGTCAA 840
QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
Db 841 GCAAGTCAAAGTAATGACATTTAGTCTCTTGAACAGCTCTACAACTGCCCTTGAGT 900
QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db 901 CAACGATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCACAAATCACAGTCAAACA 960
QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db 961 GCTAGAGGTGTTGCAGTGCCACACGGAGATCATTTACCCTTCACTCCCTTACTCTCAAATG 1020
QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340
Db 1021 TCTCAATTTGGAAGAACGAATCGCTGATTTATTTCCCTTCTGTTATCGTTCAACCATTTG 1080
QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db 1081 GTACCAGATTCAAGGCCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGGC 1140
QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
Db 1141 CCGCAACTCGACCAAAATCTTAAATAGACTCAAAATTTCTTTTGGTTAGTCACTGGTA 1200
QY 381 ArgLysValGlyGlyGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
Db 1201 CGAAAAGTTGGGGAAGGATATGTTATCGAAGAAAGGGCATCTCTCGTTATGCTTTTGGC 1260
QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
Db 1261 AAAGATTACCATCTGAAACTGTTAAATATCTTTGAAAGCAAGTTATCAAAACAAAGAGT 1320
QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
Db 1321 GTTTCACACTTTACTGCTTAAAGAAAGAAATGTTGCTCTCTGACCAAGAAATTTAT 1380
QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db 1381 GATAAAGCATATAATCTGTTAACTGAGGCTATAAAGCCCTTGTGTAATAGGGTCGT 1440
QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
Db 1441 AATTCGATTTCAGAGCTTAGACAAATTTATTAAGACGCTTGAATGATGATGATGATTA 1500
QY 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db 1501 AAAGAAAATTTGATGATGATTTTATTTGGCATTTCTTAGCACCAATTTCCCATCCAGAGCGA 1560
QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
Db 1561 CTTGGCAAAACCAATCTCAAATTTAGTACTACTGAAGACGAAGTTCGTATTTGCTCAATTA 1620
QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
Db 1621 GCTGATAGGTATACAAACGTCAGATGGTTTACATTTTGTATGAACATGATATATCACTGAT 1680
QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560
Db 1681 GAAGGAGATGCATATGTAACGCTCATATGGGCCCATAGTCACTGGGATTTGAAAGATAGC 1740

```

```

; NAME/KEY: misc feature
; LOCATION: (1)-(2451)
; OTHER INFORMATION: n = a, c, t or g
US-10-387-783-9

Alignment Scores:
Pred. No.: 0 Length: 2451
Score: 4163.00 Matches: 796
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.95% Indels: 0
DB: 16 Gaps: 0

US-09-765-271-56 (1-796) x US-10-387-783-9 (1-2451)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnArgValSerTyr 20
DB 61 TCTTACGAGTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTCCTAT 120
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
DB 121 ATAGATGGAAACCAAGCGCGCAAAAACGGAGAAATTTGACTCTCTGATGAGGTAGCAAG 180
QY 41 ArgGluGlyLysAlaGlnLysValLysLysLysLysLysLysLysLysLysLysLys 60
DB 181 CGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCACT 240
QY 61 SerHisGlyAspHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
DB 241 TCACATGGCGACCATCATATTATACATGTTAGTGTTCCTTATGACGCTATCATCAT 300
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLysLysLysLysLysLysLysLys 100
DB 301 GAAGAAATTAATCATGAAAGATCAAACTATAAGCTAAAGATGAGGATATTGTTAATGAG 360
QY 101 ValLysGlyGlyTyrValLysValLysValLysValLysValLysValLysValLys 120
DB 361 GTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGCT 420
QY 121 AlaHisAlaAspAsnValArgThrLysGluGluLysLysLysLysLysLysLys 140
DB 421 GCCACGCGGATTAACGTCCTGTAACAAAGAGGAAATCAATCGACAAAACCAAGACATAGT 480
QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
DB 481 CAACATCGTGAAGGTGGAATCCAGAAACGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 161 GlyArgTyrThrThrAspAspGlyTyrLysPheAsnAlaSerAspLysLysLysLys 180
DB 541 GGACGCTATACATACAGATGCTTATATCTTTAATGCTTCTGATATCATAGGATGAT 600
QY 181 GlyAspAlaTyrTyrLysValProHisGlyAspHisTyrHisTyrLysProLysAsnGluLeu 200
DB 601 GGTGATGCTTATATCTGTTCTCATGAGATCATCATCATCATCATCATCATCATCATCAT 660
QY 201 SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
DB 661 TCAGCTAGCGAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
DB 721 TCAAGAACCTTATCGCGCAAAATAGCATTAACCTTCAAGAACCAAACTGGGTACCTTCT 780
QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
DB 781 GTAAGCAATCCAGGAATCAAAATATAACACAAAGCAACCAAGCAACCAAGCAAGTCAA 840
QY 261 AlaSerGlnSerAsnAspLysLeuSerLysGlnLysLysLysLysLysLysLysLys 280
DB 841 GCAAGTCAAAAGTAATGACATTTAGTCTCTTGAACACAGCTCTCAAAACGCTTTGAGT 900
QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnLysThrSerArgThr 300

```

RESULT 4

```

US-10-387-783-9
; Sequence 9, Application US/10387783
; Publication No. US20040005331A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 9
; LENGTH: 2451
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:

```

Db 901 CAAAGCATGTAGAAATCTGATGCGCTTGTCTTGTATCCAGCAAAATCACAAGTCGAACA 960
QY 301 AlAaAgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db 961 GCTAGAGGTGTTCAGTGTGCACACGAGATCATTAACCTTCTATCCCTTACTCTCAATG 1020
QY 321 SerGluLeuGluGluArgIleAlaGlyIleProLeuArgTyrArgSerAsnHisTyr 340
Db 1021 TCTGAATTTGGAAGAACGAATCGCTGTATTAATTCCTTCCTCGTTATCGTTCAACCAATTGG 1080
QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db 1081 GTACACAGATTCAAGGCCAGAAACAACCAAGTCCACAAACCGACTCCGGAACCTAGTCCAGGC 1140
QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
Db 1141 CCGCAACCTGCACCAAAATCTTAAATAGACTCAAAATTTCTTTTGTGTAGTCAGCTGGTA 1200
QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
Db 1201 CGAAAGTTGGGGAAGGATATGTAATTCGAAGAAAAGGCACTCTCTGTTATGCTTTTGGC 1260
QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
Db 1261 AAAGATTACCATCTGAACCTGTAAATAATCTTGAAGCAAGATTATCAAAAACAAGAGAGT 1320
QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
Db 1321 GTTTCACACACTTAACTCTAAAAGAAAATGTTGCTCTCTGACCAAGAAATTTTAT 1380
QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db 1381 GATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATAAAGGTCGT 1440
QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
Db 1441 AATCTGATTCCCAAGCCTTAGACAAATATTATAGACGCTTGAATGATGATCGACTAAT 1500
QY 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db 1501 AAAGAAAATTGGTAGATGATTATTGTCATCTCTAGCACCAATATCCCATCCAGACGGA 1560
QY 501 LeuGlyLysProAsnSerGlnLeuGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
Db 1561 CTTGGCAACCAAAATCTCAAAATGAGTATCTGAAGCAAGTTCGTATTGCTCAATTA 1620
QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleLeuSerAsp 540
Db 1621 GCTGATAAGTATACAGCTCAGATGGTTACATTTTGTATGAACATGATATAATCAGTGAT 1680
QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560
Db 1681 GAAGGAGATGATATGTAAGCCTCATATGGGCCATAGTCACGTGGATTGGAAGATAGC 1740
QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
Db 1741 CTTTCTGATAAGGAAAAGTTGCAAGCTCAAGCTATCTAAGAAAAGATCTCTCTACCT 1800
QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600
Db 1801 CCATCTCCAGACGACATGTTAAAGCAATCCAACTGGAGATAGTCAGCAGCTATTTAC 1860
QY 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
Db 1861 AATCGTGTGAAGGGGAAAACGAATTCACCTCGTTCGACTTCCATATATGGTTGAGCAT 1920
QY 621 ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle 640
Db 1921 ACAGTTGAGGTTAAAACGGTAATTTGAATTTCTCTCAAGATCATACCAATAATATT 1980
QY 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660
Db 1981 AAATTGCTGTTGTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGAAGAT 2040

QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
Db 2041 TTGTTTGGCAGCATTAAGTACTACGTAGAACACCCCTGACGAACGTCCTCATCTTAATGAT 2100
QY 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
Db 2101 CGATGGGCAATGCGCAGTGCATGTGTAGCAAGAAAGACACACAGTGAAGTCAAT 2160
QY 701 LysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720
Db 2161 AGAACTTCAAGCCGATGAAGCCAGTAGAGGAAACACCTGCTGAGCAGAGTCCCT 2220
QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
Db 2221 CAAAGTAGAGACTGAAAAAGTAGAACCCCACTCAAGAGAGCAGAAAGTTTGTCTGCGAAA 2280
QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
Db 2281 GTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAACTCTAGCTGTTTACGAAATAT 2340
QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuAla 780
Db 2341 TTGACTCTTCAAAATTATGGATAACAATAGTATCATGGCAGAGCAGAAAAATTACTTGGC 2400
QY 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796
Db 2401 TTGTTAAAGGAAGTAACTCTTCTATCTGTAAGTAAGGAAAAATAAAC 2448

RESULT 5

US-10-158-844-94

; Sequence 94, Application US/10158844

; Publication No. US20040029118A1

; GENERAL INFORMATION:

; APPLICANT: Kunsch et al.

; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 391

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD-R

; COMPUTER: Dell Latitude Pentium 3

; OPERATING SYSTEM: Windows 98

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/158,844

; FILING DATE: 03-Jun-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/961,527

; FILING DATE: 1997-10-30

; APPLICATION NUMBER: US 60/029,960

; FILING DATE: 1996-10-31

; ATTORNEY/AGENT INFORMATION:

; NAME: Hyman, Mark J.

; REGISTRATION NUMBER: 46,789

; REFERENCE/DOCKET NUMBER: PB340PID1

; INFORMATION FOR SEQ ID NO: 94:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 8195 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 94:

US-10-158-844-94

Alignment Scores:

Pred. No.:

0

Length:

8195

Score:	4163.00	Matches:	795
Percent Similarity:	99.87%	Conservative:	0
Best Local Similarity:	99.87%	Mismatches:	1
Query Match:	99.95%	Indels:	0
DB:	13	Gaps:	0
US-09-765-271-56 (1-796) x US-10-158-844-94 (1-8195)			
QY	1	SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnArgValSerTyr	20
Db	3054	TCITACGAGTTGGACGTCTATCAAGCTAGAACGGTTAAGGAAATAAATCGTGTTCCTAT	3113
QY	21	IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys	40
Db	3114	ATAGATGMAAACCAAGCGACGCAAAACCGGAGAAATTTGACTCCTCGATGAGGTAGCAAG	3173
QY	41	ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr	60
Db	3174	CGTAGGAAGTCAATGCTGAGCAATCGTCATCAAGATAACAGACCAAGGCTATGTCAT	3233
QY	61	SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer	80
Db	3234	TCACATGGCGACCATCATATTATCAATGTAAGGTTCCCTATGACGCTATCATCAT	3293
QY	81	GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu	100
Db	3294	GAAGAATTTACTCATGAAAGATCCAACTATAAGCTAAAGATGAGGATATTGTTAATGAG	3353
QY	101	ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla	120
Db	3354	GTCAAGGTGGATATGTTATCAAGTAGTAGTGAATAATACTANGTTTACCTTAAGATGCT	3413
QY	121	AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer	140
Db	3414	GCCACGCGGATAACGTCGGTACAAAGAGGAATCAATCGCAAAACCAAGAGCATAGT	3473
QY	141	GlnHisArgGluGlyThrProArgAsnAspGlyValAlaValAlaAlaArgSerGln	160
Db	3474	CAACATCGTGAAGGTGGAACTCCAGAAACGATGGTGTCTTCCCTGGCACGTTGCGAA	3533
QY	161	GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr	180
Db	3534	GGACGCTATCTACATGATGTTATATCTTAAATGCTTCTGATATCATAGAGGATACT	3593
QY	181	GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu	200
Db	3594	GGTGAATGCTTATATCGTCTCTCATGAGATCAATCAATACATTCCTAAGAAATGATTA	3653
QY	201	SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn	220
Db	3654	TCAGCTAGCCAGTTGGCTGCTGCAAGAGCCCTTCTATCTGGTCAGGAAATCTGTCBAAT	3713
QY	221	SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer	240
Db	3714	TCAGAACCTATCCCGCAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTCT	3773
QY	241	ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln	260
Db	3774	GTAAGCAATCCAGAACTACAAATACTAACACAAAGCAACACACACACTAACAGTCAA	3833
QY	261	AlaSerGlnSerAsnAspIleAspSerLeuLysGlnLeuTyrLysLeuProLeuSer	280
Db	3834	GCAAGTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	3893
QY	281	GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr	300
Db	3894	CAACGACATGTAGATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	3953
QY	301	AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet	320
Db	3954	GCTAGAGGTGTTCAGTGGCCACGAGATCATACCACTTCACTTCACTTCACTTCACTTCA	4013
QY	321	SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp	340

Db	4014	TCGAATTTGAAGAACGAATCGCTCGTATTATTTCCCTTCCTGTTATCGTTCAACCAATGG	4073
QY	341	ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly	360
Db	4074	GTACAGATTCAAGCCGACGAACCAACCAAGTCCACCAACCGACTCCGGAACCTAGTCCAGC	4133
QY	361	ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal	380
Db	4134	CCGCAACCTGCACCAATCTTAAATAGACTCAAAATCTCTTTGGTTAGTCAGCTGGTA	4193
QY	381	ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla	400
Db	4194	CGAAAGTTGGGGAAGGATATGTTTGAAGAAAGGCACTCTCCGTATGTCTTTGCG	4253
QY	401	LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer	420
Db	4254	AAAGATTTTACCATCTGAACCTGTTAAATCTTGAAGCAAGTTATCAAAACCAAGAGAT	4313
QY	421	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	440
Db	4314	GTTCACACACTTTAACTGTCTAAAGAAAGAAATGTTCTCTCGTGACCAAGAAATTTAT	4373
QY	441	AspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg	460
Db	4374	GATAAAGCATATAATCTGTAACTGAGGCTCATAAAGCTTGTGTTGAAAATAAGGCTCGT	4433
QY	461	AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn	480
Db	4434	AAATCTGATTTCCAGACCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAT	4493
QY	481	LysGluLysLeuValAspAspLeuAlaPheLeuAlaProIleThrHisProGluArg	500
Db	4494	AAAGAAAAATTTGGTATGATTTATTTGCAATTTCTTAGCACCAATTACCCATCCAGACCA	4553
QY	501	LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu	520
Db	4554	CTTGGCAAACTTCTCAAAATTTGATATCTAGAGCAAGTTCGTATGCTCAATTA	4613
QY	521	AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleSerAsp	540
Db	4614	GCTGATAGTATACACGTCAGATGTTTACATTTTGTGATGACATGATTAATCACTGAT	4673
QY	541	GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer	560
Db	4674	GAAGAGATGCATATGATTAACGCTCATATGGCCATAGTCAGTGGATGGAAAGATAGC	4733
QY	561	LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro	580
Db	4734	CTTCTGTAAGGAAAGTTGCGCTCAAGCTTACTTAAGCAAGGATATACCTACCT	4793
QY	581	ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr	600
Db	4794	CCATCTCCAGACCGAGATGTTAAAGCAATCCAACTGGAGATAGTCAGCAGCTATTATAC	4853
QY	601	AsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGluHis	620
Db	4854	AAATCGTGAAGGGAAGGAAACGAATTCCTCGTTCCTGATTCATATATGTTGACAT	4913
QY	621	ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle	640
Db	4914	ACAGTTGAGGTTTAAACCGGTAAATTTGATTTCTCTCATAGGATCATTAACCAATATTT	4973
QY	641	LysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp	660
Db	4974	AAATTTGCTGTTGATGATCAACATCAAAAGCTCCAAATGGCTATACCTTGGAGAT	5033
QY	661	LeuPheAlaThrIleLysTyrValGluHisProAspGluArgProHisSerAsnAsp	680
Db	5034	TTGTTGCGACGATTAAGTACTAGTAGAACCCCTCGACACCTGCCACATTTCAATGAT	5093
QY	681	GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn	700


```

Db      5094 GGATGGGCAATGCCAGTGGATGTTAGGCAAGAAAGACCAAGTGAAGATCCAAAT 5153
Qy      701  LysAsnPheLysAlaAspGluProValGluGluThrProAlaGluProGluValPro 720
Db      5154 AAGAATCTCAAGCGATGAAGAGCGAGTAGAGGAAACACCTGCTGAGCCAGAGTCCCT 5213
Qy      721  GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
Db      5214 CAAAGTAGAGTGAAGAAAGTAGAAGCCCAACTCAAAGAGCAGAGATTGCTTGGGAA 5273
Qy      741  ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
Db      5274 GTAACGGATTCTAGTCTCAAGCCCAATGCAACAGAAACTCTAGCTGGTTACGAAATAAT 5333
Qy      761  LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780
Db      5334 TTGACTCTTCAAAATTATCGATAACATAGTATCATGGCAGAGCAGAAATAATTACTGGC 5393
Qy      781  LeuLeuLysGlySerAsnProSerValSerLysGluLysIleAsn 796
Db      5394 TTGTTAAAGGAAGTAATCTTCTCATCTGTAGTAAGGAAATAAATAAAC 5441

RESULT 6
; Sequence 3, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 2523
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; US-09-884-465A-3

Alignment Scores:
Pred. No.: 2,37e-282 Length: 2523
Score: 3218.00 Matches: 616
Percent Similarity: 82.95% Conservative: 65
Best Local Similarity: 75.03% Mismatches: 102
Query Match: 77.26% Indels: 38
DB: 10 Gaps: 3

US-09-765-271-56 (1-796) x US-09-884-465A-3 (1-2523)

Qy      1  SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db      61  GCTTATGAATAGTTGATCGATCAAGCTCAACTGTAAAGAAATAATCGTGTTCCTAT 120
Qy      21  IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db      121  ATAGATGGAAAAACAAGCGCAAAAAACGAGAAATTTGACTCTCTGATGAGGTAGCAAG 180
Qy      41  ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db      181  CGTGAAGGAATCAACCCGCAACAAATCGTCATCAAGATTACGATCAAGTTATGTGACC 240
Qy      61  SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db      241  TCTCATGGAGACCATTTACTTACTATAATGGCAAGGTCCTTATGATGCCATCATCAGT 300

```

```

Qy      81  GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db      301  GAAGAGCTCTCATGAAGATCCGAATTATCAGTTGAAGATTTCAGACATTGTCAATGAA 360
Qy      101  ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db      361  ATCAAGGTGGTGTATGTCAATTAAGGTAAACGGTAAATACTATGTTTACCTTAAGGATGA 420
Qy      121  AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
Db      421  GCTCATCGGATAATGTCCGTACAAAGAGAAATCAATCGCAAAACACAGAACATAGT 480
Qy      141  GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
Db      481  CAGCATCTGTGAAGGAGGACTTCAGCAAAACGATGTCGGTAGCCTTTGCGACGTTCCACAG 540
Qy      161  GlyArgTyrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db      541  GGACGCTACACACAGATGATGTTTATCTTCAATGATCATCTGATATCATCGAAGATAGC 600
Qy      181  GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db      601  GGCGATGCTTATATCGTTCTCATGGAGATCATACCATTTACATTCCTTAAGATGAGTTA 660
Qy      201  SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db      661  TCAGTAGCGAGTTGGCTGTCAGAAAGCCTTCCTATCTGCTCGGAAAAATCTGTCAAAAT 720
Qy      221  SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
Db      721  TTAAGAACCTTATCGCCGACAAATAGCGATAACACTCCAAGAACAAACTGGGTACTTCT 780
Qy      241  ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db      781  GTAAGCAATCCAGGAATCTACAAATACTAACACAGCAACACACACACACTTAACGTCAA 840
Qy      261  AlaSerGlnSerAsnAspIleAspSerLeuLysGlnLeuLysGlnLeuLysLeuProLeuSer 280
Db      841  GCAAGTCAAAAGTAATGACATTGATGATCTCTTGAACACAGCTCTACAAACTGSCCTTTGAGT 900
Qy      281  GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db      901  CAACGCCATGTAGATCTGATGGCTTTATTTTCGACCCAGCGCAAAATCACAAAGTCGAACC 960
Qy      301  AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db      961  GCCAGAGGTGATGCTGTCCTCATGTGTAACCATTTACCACCTTATCCCTTTATGAACAAATG 1020
Qy      321  SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340
Db      1021  TCTGAATTTGGAAAAACGAATTTGCTCGTATTATTTCCCTTCGTTATGTTTCAAACTTGG 1080
Qy      341  ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db      1081  GTACACAGATTCAAGACCAAGAACCAAGTCCAAACCGACTCCAGAACCTAGTCCCAAGT 1140
Qy      361  ProGlnProAlaProAsnLeuLys-----IleAspSerAsnSerSer 374
Db      1141  CGCAACCTTCACCAATCTCAACCCAGCTCCCAAGCAATCCAATTGAT-----GAGAAA 1194
Qy      375  LeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIle 394
Db      1195  TTGGTCAAAAGAGCTGTTCGAAAAAGTAGCGATGTTATGTCTTTGAGGAGATGAGATT 1254
Qy      395  SerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLys 414
Db      1255  TTTCTGTTATATCCCGCCCAAGAAATCTTTCAGCAGAAACAGCAGCAGCGCATTCATGACAAA 1314
Qy      415  LeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysGluAsnValAlaPro 434
Db      1315  CTGGCCACAGCAGAAAGTTTATCTCATAGCTAGAGCTTAAGAAACGACCTCCCATCT 1374
Qy      435  ArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeu 454

```

Db 1375 AGTGATCGAAGATTTTCAATTAAGGCTTATGACTTACGACAAAGTTCACCAAGATTTA 1434
 QY 455 Phe**AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeu 474
 Db 1435 CTTGATATATAAGGTCGACAAAGTTGATTTTGAGGCTTTGGATAACCTGTGTGAACGACTC 1494
 QY 475 AsnAspGluSerThrAsnLysGluLysLeuValAspLeuLeuAlaPheLeuAlapro 494
 Db 1495 AAGGATGCTCAAGTGATAAAGTCAAGTTAGTGATGATATCTTGCCCTTTTAGCTCCG 1554
 QY 495 IleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluLysThrGluAspGlu 514
 Db 1555 ATTCTATCAGAACGCTTTAGAAACCAATATGCGCAATATCTACACATGATGATGAG 1614
 QY 515 ValArgIleAlaGlnLeuAlaAspLysThrThrSerAspGlyTyrIlePheAspGlu 534
 Db 1615 ATTCAAGTAGCAAGTTGCGAGGCAAGTACACACAGAAAGCGGTATATCTTTGATCCT 1674
 QY 535 HisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHis 554
 Db 1675 COTGATATAACCAAGTATGAGGGGATGCTATGTAATGTAATGTAATGTAATGTAATG 1734
 QY 555 TrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLys 574
 Db 1735 TGGAITATAAAGATAGTTGTTCTGAAGCTGAGAGCGCGACGCCAGCTTATGCTAAA 1794
 QY 575 GluLysGlyIleLeuProSerProAspAlaAspValLysAlaAsnProThrGlyAsp 594
 Db 1795 GAGAAAGGTTTGACCCCTCTTCGACAGACCATCAGGATTCAGAAATACTGAGGCAAAA 1854
 QY 595 SerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArgLeu 614
 Db 1855 GGAGCAGAAAGCTATCTACAAACCGCTGAAAGCAGCTTAAGAGGTCCTTGTATGATG 1914
 QY 615 ProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHisLys 634
 Db 1915 CTTTACAATCTTCAATATACTGTAGAGTCAAACCGGTAGTTTAATCATACCTCATTAT 1974
 QY 635 AspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsn 654
 Db 1975 GACCATTAACATAACATCAAAATTTGAGTGTGTTTGACGAAGGCTTTTATGAGCACCCTAAG 2034
 QY 655 GlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGlu 674
 Db 2035 GGGTATACTCTTGAGGATCTTTGGGACTGTCTAGTACTGTAAGTCTGAAATCCAAACGAA 2094
 QY 675 ArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAsp 694
 Db 2095 CGTCCGCATTCAGATAATGTTTGGTAAAGCTAGCGACCATGTTCAAAAGAAACAAAAT 2154
 QY 695 HisSerGluAspProAsnLysAsnPheLysAlaAsp----- 706
 Db 2155 GGTCAAGCTGATACCAATCAACCGGAAACCAACGAGGAGGAGAAACCTCAGACAGAAAA 2214
 QY 706 ----- 706
 Db 2215 CTTGAGGAAGAAACCCCTCGAAGAGAAACCAACAGGAGGAGAAACAGAGTCTCCAAAA 2274
 QY 707 -----GluGluProValGluGluThrProAlaGluProGluValProGluValGluThr 724
 Db 2275 CCACACAGAGAACCCAGAGAGAAATCACCAGAGGAATCAGAGAACTCAGGTCGAGACT 2334
 QY 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer 744
 Db 2335 GAAAGGTTGAAGAAAAAAGCTGAGAGAGGCTGAGATTTACTTGAAAAATCCAGAGTCCA 2394
 QY 745 SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnLeuThrLeuGln 764
 Db 2395 ATTATCAAGTCCCAATGCGCAAGAGACTCTCAGAGATTAAATAATTTACTATTGGC 2454
 QY 765 IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLeuLysGly 784

Db 2455 ACCGAGCAACAATACTATTATGGCAGAAAGCTGAAAAAACTATTGGCTTTTATTAAAGGAG 2514
 QY 785 Ser 785
 Db 2515 AGT 2517
 RESULT 7
 US-09-884-465A-4
 ; Sequence 4, Application US/09884465A
 ; Publication No. US20030077293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis
 ; APPLICANT: Charland, Nathalie
 ; APPLICANT: Ouellet, Catherine
 ; TITLE OF INVENTION: Streptococcus Antigens
 ; FILE REFERENCE: 055190-0044
 ; CURRENT APPLICATION NUMBER: US/09/884,465A
 ; CURRENT FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: 60/212,683
 ; PRIOR FILING DATE: 2000-06-20
 ; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 4
 ; LENGTH: 2647
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 ; US-09-884-465A-4
 Alignment Scores:
 Pred. No.: 2,53e-282 Length: 2647
 Score: 3218.00 Matches: 616
 Percent Similarity: 82.95% Conservative: 65
 Best Local Similarity: 75.03% Mismatches: 102
 Query Match: 77.28% Indels: 38
 DB: 10 Gaps: 3
 US-09-765-271-56 (1-796) x US-09-884-465A-4 (1-2647)
 QY 1 SerTyrGluLeuGlyLysLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
 Db 105 GCTTATGAATAGGTTTGCATCAAGCTCAATCTGTAAGAAATATCGTGTTCCTAT 164
 QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
 Db 165 ATAGATGGAACAAACAGCAGCAGCAAAACCGAGAAATTTGACTCCTCATGAGGTAGCAAG 224
 QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
 Db 225 CGTGAAGGAATCAACGCCCAACAAATCGTCATCAAGATTACGGATCAAGGTATGTGACC 284
 QY 61 SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
 Db 285 TCTCATGGAGACATTTATCTACTATATATGGCAAGGTCCCTTATGATGCCATCATCAGT 344
 QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
 Db 345 GAAGAGCTCTCATGAAGATCGAATTTATCAGTTCAAGGATTCAGACATTTGTCAATGAA 404
 QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrValLysAspAla 120
 Db 405 ATCAAGGGTGGTTATGTCTATTAAGGTAACCGGTAATAATCTATGTTTACCTTAAAGGATGCA 464
 QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
 Db 465 GCTCATGGGATATGTCCGTACAAAGAGAAATCAATCGCAAAACCAAGAACATAGT 524
 QY 141 GlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
 Db 525 CAGCATGCTGAAGGAGGACTTCAGCAAAACGATGTCGGTTCGGTAGCTTTGCACGTTTCAG 584

QY 161 GlyArgTyrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
 DB 585 CGACGCTACACACAGATGATGTTATATCTTCAATGCATCTGATATCATCGAAGTACG 644
 QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProHisAsnGluLeu 200
 DB 645 GCGCATGCTATATCGTTCTCATGAGATCATATACCATATACATTCCTAAGAATGAGTTA 704
 QY 201 SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
 DB 705 TCAGCTAGCGAGTTGGCTGCTCAGAGCCCTTCTATCTGGTCGGGAAATCTGCAAAAT 764
 QY 221 SerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
 DB 765 TTAAGAACCTATCGCCGCAAAATAGCGATAACACTCCAAGAACAACCTGGGTACTTCT 824
 QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
 DB 825 GTAAGCAATCCAGGAACCTACAAATACTAACCAAGCAACACAGCAACACTAACAGTCAA 884
 QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLeuGlnLeuTyrLysLeuProLeuSer 280
 DB 885 GCAAGTCAAGTAATGATGATGATCTCTTGAAACAGCTCTCAAACTGCTTTGAGT 944
 QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
 DB 945 CAACGCCATGTAGATCTGATGGCTTATTTTCGACCCAGCGCAATCAAGTCGAACC 1004
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
 DB 1005 GCCAGAGTGTAGTGTCTCCCTCATGTTAACCATTACCACTTATTCCTTATGAACAAATG 1064
 QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340
 DB 1065 TCTCAATTGGAAACGAATGCTCGTATATTCCTTCCTTCCTTATCGTTCAAAACCATTTG 1124
 QY 341 ValProSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
 DB 1125 GTACCAATTCACAGCAGAGAACCAAGTCCACACCGACTCCAGAACCTAGTCCAACT 1184
 QY 361 ProGlnProAlaProAsnLeuLys-----IleAspSerAsnSerSer 374
 DB 1185 CCGCAACCTGCACCAATCTCTCAACCAAGCTCCAAACCAATCAATTTGAT-----GAGAAA 1238
 QY 375 LeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIle 394
 DB 1239 TTGGTCAAGAGAGCTGTTCGAAAGTAGGAGTGTATGCTTTTGGAGGAGATGGAGTT 1298
 QY 395 SerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLys 414
 DB 1299 TCTCGTTATATCCCGCAAGAACTTTTCAGCAGAAACACAGCAGCGCATTTGATAGCAA 1358
 QY 415 LeuSerLysGlnGluSerValSerHisThrIleuThrAlaLysLysGluAsnValAlaPro 434
 DB 1359 CTGCGCAAGCAGGAAAGTTTATCTCATAGCTAGGAGCTAAGAAAACCTGACCTCCCATCT 1418
 QY 435 ArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeu 454
 DB 1419 AGTGATCGAGAAATTTACATTAAGCTTAGCTTATGACTTACCAAGAAATTCACCAAGATT 1478
 QY 455 Phe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuArgLeu 474
 DB 1479 CTTGATAATAAAGTCGACAAAGTTGATTTTGAGCTTTGGATTAACCTTTGGAACCACTC 1538
 QY 475 AsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaPro 494
 DB 1539 AAGGATGCTCAAGTATAAAGTCAAGTTAGTGGATGATATCTTCTGCTTCTTAGCTCCG 1598
 QY 495 IleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGlu 514
 DB 1599 ATTCTGTCATCCGAAAGTTTAGGAAACCAATTCGCCAAATACCTACACTGATGATGAG 1658
 QY 515 ValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGlu 534

DB 1659 ATTCAAGTAGCCAAAGTTGGCAGCGCAAGTACACAACAGAGAGCGGTATATCTTTGATCT 1718
 QY 535 HisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHis 554
 DB 1719 CQTGATATAACAGTGATGAGGGGATGCTATGTAACCTCCACATATACCATAGCCAC 1778
 QY 555 TrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLys 574
 DB 1779 TGGATTAATAAAGATAGTTTGTCTGAAGCTGAGAGCGGCGCAGCCAGGCTTATGCTAAA 1838
 QY 575 GluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGlyAsp 594
 DB 1839 GAGAAAGGTTTGACCCCTCTTCGACAGACCATCAGGATTCAGGAAATACTCAGGCAAAA 1898
 QY 595 SerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArgLeu 614
 DB 1899 GGAGCAGAAGCTATCTACAAACGCTGAAGCAGTAAGAGGTGCCACTTGTATCGTATG 1958
 QY 615 ProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHisLys 634
 DB 1959 CTTTACCAATCTCAATATCTAGTAAAGTCAAAAACGCTAGTGTATATCATCTCATAT 2018
 QY 635 AspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsn 654
 DB 2019 GACCATTACCATAACATCAAAATTTGAGTGGTTTTCGAAAGGCTTTATCAGSCACCTAAG 2078
 QY 655 GlyTyrThrLeuGluAspPheAlaThrIleLysTyrValGluHisProAspGlu 674
 DB 2079 GGTATATCTCTTGGAGTCTTTTGGCCAGCTCTCAAGTACTATGTCGAACATCCCAACGAA 2138
 QY 675 ArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAsp 694
 DB 2139 CGTCCGATTCAGATAATGTTTGTGTAACGCTAGCGACCATGTTCAAAGAAACAAATAAT 2198
 QY 695 HisSerGluAspProAsnLysAsnPheLysAlaAsp----- 706
 DB 2199 GGTCAAGCTGATACCAATCAAAACGGAACCAACGAGGAGGAGAAACCTCAGACAGAAAA 2258
 QY 706 ----- 706
 DB 2259 CCTGAGCAAGAAACCCCTCGAGAGAGAAACCAACAGCGAGAAACCCAGAGCTCTCCAAAA 2318
 QY 707 -----GluGluProValGluGluThrProAlaGluProGluValProGlnValGluThr 724
 DB 2319 CCAACAGAGAAACCAAGAAAGAAATCACCAGAGGAATCAGAGAACCTCAGTTCGAGACT 2378
 QY 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer 744
 DB 2379 GAAAAGGTTGAAGAAACCACTGAGAGAGGCTGAAGATTTTACTTGGAAAAATCCAGGATCCA 2438
 QY 745 SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnLeuThrLeuGln 764
 DB 2439 ATTATCAAGTCCCAATGCCAAAGAGACTCTCAGAGGATTAATAAATAATTTACTATTGGC 2498
 QY 765 IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGly 784
 DB 2499 ACCCAGGACCAACAATACTATATGCGCAGAGCTGAAGAACTATTTGGCTTTATTAAAGGAG 2558
 QY 785 Ser 785
 DB 2559 AGT 2561

RESULT 8

; US-09-769-787-206
 ; Sequence 206, Application US/09769787
 ; Publication No. US20030091577A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Microbial Technics Limited
 ; APPLICANT: Gilbert, Christophe FG
 ; APPLICANT: Hansbro, Philip M
 ; TITLE OF INVENTION: proteins
 ; FILE REFERENCE: PWC/P21129W0

CURRENT APPLICATION NUMBER: US/09/769,787
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: GB 9816337.1
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: US 60/125164
PRIOR FILING DATE: 1999-03-19
NUMBER OF SEQ ID NOS: 388
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 206
LENGTH: 2481
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-769-787-206

Alignment Scores:
Pred. No.: 6,58e-244 Length: 2481
Score: 2795.00 Matches: 549
Percent Similarity: 76.98% Conservative: 83
Best Local Similarity: 66.87% Mismatches: 137
Query Match: 67.11% Indels: 52
DB: 10 Gaps: 8

US-09-765-271-56 (1-796) x US-09-769-787-206 (1-2481)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnArgValSer 19
DB 61 TCCTATGAGCTGGTGGTCGTCACCAAGCTGGTCAGGATAGAAAGAGTCTATACGATTGCT 120
QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
DB 121 TATATAGATGGTGATCAGCTGGTCAAAGGCGAGAAACTTGACACCAGATCAAGTCAGT 180
QY 40 LysArgGluGlyLeuAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
DB 181 AAGAGGAGGGGATCAACGCCGAAACAAATCGTCATCAAGATTACGGATCAAGGTTATG 240
QY 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIle 79
DB 241 ACCTCTCATGGAGACCATATCATCTATTAATGGCAAGGTCCTTATGATGCCATATC 300
QY 80 SerGluGluLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
DB 301 AGTGAGAGCTCTCATGAAGATCGGAATTCAGTTGAAGGATTACAGACATTTGCAAT 360
QY 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysAsp 119
DB 361 GAAATCAAGGGTGGTATGTGTATCATCAAGGTAGACGGAAATATCTATGTTACCTTAAGGAT 420
QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
DB 421 GCAGCTCATGCGGATATATTCGACAAAGAGAGATTAACGTCAGAGCAGGAACGC 480
QY 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
DB 481 AGTCATAATCAC-----GGTCCAGGAGCTAACGATCATGCTAGCTGAGCCAGAGCC 534
QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
DB 535 CAAGGAGCGCTATACACGGATGATGGGTATATCTTCAATGCATCTGATATCATITAGGAC 594
QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
DB 595 ACGGGTGATGCTTATATCTGCTTCCTCAGCGGACCATATACCATTCCTAAGAATGAG 654
QY 200 LeuSerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
DB 655 TTATCAGCTAGCAGATTAGCTGCTGCAGAGCCTATTGGAATGGG----- 699
QY 220 AsnSerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239
DB 700 -----AAGCAGGGATCTCGCTCTTCAAGTCTTAGTTATATGCA 741
QY 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259

DB 742 AATCCAGCTCAACCAAGATTGTTCAGACACCAAACTGTGACTGTCTCACTCCAACTTATCAT 801
QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
DB 802 CAA--AATCAAGGGGAAACATTTCAAGCTTTTACGTGAATGTATGCTAAACCCCTTA 858
QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
DB 859 TCAGAACGCCATGTGAATCTGATGGCTTATTTTCGACCCAGCGCAATCACAAGTCGA 918
QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
DB 919 ACCGCCAGAGGTAGCTGTCCCTCATGGTAACCATACCACTTTATCCCTTATGAAACA 978
QY 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
DB 979 ATGCTGAATTTGGAAAAAGAAATGCTCGTATTTATCCCTTCGTTATCGTTCAACCAT 1038
QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
DB 1039 TGGGTACAGATTCAAGACCAAGCAACCAAGTCCACAATCGACTCCGGAACCTAGTCCA 1098
QY 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373
DB 1099 AGTCGCAACCTGCAACCAATCTCAACCAGCTCCAAAGCAATCCAAATGAT-----GAG 1152
QY 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393
DB 1153 AAAATGGTCAAGAAAGCTGTTCCAAAAGTAGGCGATGTTATGTTCTTTGAGGAGAATGCA 1212
QY 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413
DB 1213 GTTTCCTGTTATATCCAGCAAGGATCTTTCAGCAGAAACAGCAGCAGCATTTGATAGC 1272
QY 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433
DB 1273 AAAATGGCCCAAGCAGGAAAGTTATCTCATAGCTAGGAGCTAAGAAACTACCTCCCA 1332
QY 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453
DB 1333 TCTAGTATCGAAGATTTTACAAAGGCTTATGACTTACTAGCAAGAATTCACCAAGAT 1392
QY 454 LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473
DB 1393 TTACTTTGATAATAAAGGTCGACAAAGTTGATTTTGGAGCTTTGGATAACCTGTGGAACGA 1452
QY 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla 493
DB 1453 CTAAGGATGTCCCAAGTCAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1512
QY 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513
DB 1513 CCGATTCTGTCATCCAGACGTTTAGGAAAAACCAATGCGCAATTACCTACACTGATGAT 1572
QY 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533
DB 1573 GAGATTCAAGTACCAAGTTGGCAGGCAAGTACACAAACAGAGACCGGTATATCTTTGAT 1632
QY 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553
DB 1633 CCTCGTGATATACCGATGATGAGGGGATGCTTATGTAATCTCCATATGAGCCCATAGC 1692
QY 554 HistripleGlyLysAspSerLeuSerLysGluLysValAlaGlnAlaTyrThr 573
DB 1693 CACTGGATTAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCGAGCCAGGCTTATGCT 1752
QY 574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593
DB 1753 AAAAGAGAAAGGTTTGAACCTCTCTTCAGACAGCATCAGGATTCAGGAAATCTAGGCGCA 1812
QY 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg 613

976 ATGCTGAATTGGAAACGAATTCGCTGATTAATTCCTTCGTTATCGTTCAACACCAT 1035
 340 TrpValProAspSerArgProGluGlnProSerProGluProThrProGluProSerPro 359
 1036 TGGGTACCAGATTCAAGACAGAGAACCAAGTCCACACCGACTCCAGAACCTAGTCCA 1095
 360 GlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeu 379
 1096 AGTCGGCACCATCTCCAGCAATCCCAATTGAT-----GGGAAATTGGTCAAGAAAGCT 1149
 380 ValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPhe 399
 1150 GTTCGAAAGTAGCGATGTTATGTTCTTTCGAGGAGAAATGAGTCTTCGTTATATCCCA 1209
 400 AlalysAspLeuProSerGluThrValLysAsnLeuGluSerLysLysGlnGlu 419
 1210 GCCAAGGATCTTCAGCAGAACACAGCAGCAGGCAATGATAGCAACTGGCCACAGCAGAA 1269
 420 SerValSerHisThrLeuThrAlalysLysGluAsnValAlaProArgAspGlnGluPhe 439
 1270 AGTTTATCTCATAGCTAGGAACCTAAGAAACTGACCTCCCATCTAGTGTATCGAGAATTT 1329
 440 TyrAspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGly 459
 1330 TACAATAAGGCTTATCACTTACTAGCAAGAAATCACCAGAAATTTTACTTGATATATAAGGT 1389
 460 ArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThr 479
 1390 CCACAAAGTTGATTTTGGAGCTTTGGATTAACCTGTTGGACGACTCAAGGATGTCCTCAAGT 1449
 480 AsnLysGluLysLeuValAspLeuLeuAlaPheLeuAlaProIleThrHisProGlu 499
 1450 GATAAAGTCAAGTTAGTGAAGATATCTTTCCTTCTTAGCTCGGATTCGTCATCCAGAA 1509
 500 ArgLeuGlyLysProAsnSerGlnLeuGluThrGluAspGluValArgIleAlaGln 519
 1510 CGTTTAGGAAACCAATCCGCAATATCACTACATGATGATGAGATTCAAGTAGCCAG 1569
 520 LeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleSer 539
 1570 TTGGCAGGCAGTACACAGCAGACGAGCGTTATATCTTTGATCCTCGGTGATATAACAGT 1629
 540 AspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyIleAsp 559
 1630 GATGAGGGGATGCCTATGTAACTCCACATATGACCCATAGCCACTGGATTAATAAAGAT 1689
 560 SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeu 579
 1690 AGTTTGTCTGAAGCTGAGAGCGGCGGAGCCCGCTTATGCTTGAAGAGAAAGGTTTGACC 1749
 580 ProProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaAlaIle 599
 1750 CTTCTTCGACAGCATCATCAGGATTCAGGAAATACTGAGGCAAAAGGAGCAGAGCTATC 1809
 600 TyrAsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGlu 619
 1810 TACAACCGMTGAAAGCAGCATTAAGAGGTCGCACTTGATCGTATGCCCTTACAACTTCCA 1869
 620 HisThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsn 639
 1870 TATACCTGTAAGTCAAAACCGTAGTTTAACTATCACTCATATGACCATTAACATAAC 1929
 640 IleLysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGlu 659
 1930 ATCAAAATTTGAGTGTGTTGCAAGGCGCTTATGAGGCACTTAAGGGGTATATCTCTGAG 1989
 660 AspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsn 679
 1990 GATCTTTTGGCAGCTGCTCAAGTACTATGTGCAACATCCAAACGAACTCCGATTCAGAT 2049
 680 AspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspPro 699
 ...
 2050 AATGGTTTTTGGTAACTAGCGACCATGTTCAAAGAAACAAAAATGGTCAAGCTGATACC 2109
 700 AsnLysAsnPhelysAlaAspGluGlu-----ProValGluGluThr 713
 2110 AATCAACAGGAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACCTGAGGAAGAAC 2169
 713 ----- 713
 2170 CCTCGAAGAGAAACCGCAACCGAGAAACAGAGTCTCCAAACCAACAGAGGAACCA 2229
 714 -----ProLaGluProGluValProGlnValGluThrGlnLysValGluAlaGln 730
 2230 GAAGAAATCACCAGAGGAATCAGAAGAACCTCAGTTCGAGCTGAAAGAGTTGAAGAA 2289
 731 LeuLysGluAlaGluValLeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAla 750
 2290 CTGAGAGAGGCTGAAGATTTTACTTTGGAAANAATCCAGATCCCAATATCAAGTCCATGCC 2349
 751 ThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSer 770
 2350 AAGAGAGCTCTCACAGGATTTAAAAATAATTTACTATTTGGCACCCAGGACCAACATACT 2409
 771 IleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer 785
 2410 ATTATGGCAGAGCTGAAAAACTATTGGCTTTTATTAAGGAGAGT 2454
 2410 ATTATGGCAGAGCTGAAAAACTATTGGCTTTTATTAAGGAGAGT 2454
 RESULT 10
 US-10-412-850-11
 ; Sequence 11, Application US/10412850
 ; Publication No. US20040001836A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
 ; FILE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-686
 ; CURRENT APPLICATION NUMBER: US/10/412,850
 ; CURRENT FILING DATE: 2003-04-14
 ; PRIOR APPLICATION NUMBER: 09/468,656
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 3.0
 ; SEQ ID NO 11
 ; LENGTH: 2531
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-10-412-850-11
 Alignment Scores:
 Pred. No.: 4,01e-243 Length: 2531
 Score: 2786.50 Matches: 544
 Percent Similarity: 77.06% Conservative: 84
 Best Local Similarity: 66.75% Mismatches: 140
 Query Match: 66.90% Indels: 47
 Dbs: 16 Gaps: 7
 US-09-765-271-56 (1-796) x US-10-412-850-11 (1-2531)
 Qy 1 SerTyrLeuLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnArgValSer 19
 Db 61 TCCTATGAGCTTGAGCTTACCAAGCTGTCAGAGTAAAGAGTCTAATCGAGTTGCT 120
 Qy 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
 Db 121 TATATAGATGGTGCATCAGGCTGGTCAAAAGCAGAAACTTGACACCATGAGTCACT 180
 Qy 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
 Db 181 AAGAGGGAGGGGATCAACGCCGAAACAAATTTGTTTATCAAGATTCAGGATCAAGGTTATGTG 240

QY 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIle 79
 Db 241 ACCTCTCATGGAGACCATTAATCTATGATGCAAGGTTCCCTTATGATGCCATCATC 300
 QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
 Db 301 AGTGAAGAGCTCCTCATGAAGATCCGAATTAATCATGTTGAAGGATTCAGACATTGTCAAT 360
 QY 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysLysAsp 119
 Db 361 GAAATCAAGGCTGGTATGCTAATAGGTAACCGTAATAATCTATGTTTACCTTAAGGAT 420
 QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
 Db 421 GCRGCTCATGCGGATAATATTCGGACAAAGAGAGATTAAACGTCAGAGCGAGAACGC 480
 QY 140 SerGlnHisArgGluGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
 Db 481 AGTCATAATCATCACTCA-----AGAGCAGATAATGCTGTGCTGCAGCCAGAGGCC 531
 QY 160 GlnGlyArgTyrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
 Db 532 CAAGGACGTTATACACCGATGATGGGTATATCTTCAATGCATCTGATATCATTCAGGAC 591
 QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
 Db 592 ACGGCTGATGCTTATATCGTTCCTCACGCGACCATTAACCATTAATTCCTTAAGNATGAG 651
 QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuLeuSerGlyArgGlyAsnLeuSer 219
 Db 652 TTATCAGCTAGCAGATGATGCTGCGAGAGCCATATTGGAATGG----- 696
 QY 220 AsnSerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239
 Db 697 -----AAGCAGGATCTCGTCTCTTCAAGTTCTAGTTAATATGCA 738
 QY 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
 Db 739 AATCCAGCTCAACCAAGATTGTCAGAGAACCCACAACTCTGATGTCACCTCCACTTATCAT 798
 QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
 Db 799 CAA--AATCAAGGGGAAAAATTTCAAGCCCTTTAGCTGAATGTATGCTAAACCCCTTA 855
 QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
 Db 856 TCAGAACCCCATGTGGAAATCTGATGGCTTATTTTCAGCCAGCGCAATCACAAGTCGA 915
 QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
 Db 916 ACCGCCAGAGGTGTAGTGTCCCTCATGTTAACAATTAACCATTTATCCCTTATGAACAA 975
 QY 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
 Db 976 ATGTCTGAATGGAAACCAAGATTTGCTGCTATTATTCCTTCCCTTATTCCTTCAACCAT 1035
 QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
 Db 1036 TGGGTACCATGATTCAGACCCAGAGAACCAAGTCCACACCGACTCCAGAACCTAGTCCA 1095
 QY 360 GlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeu 379
 Db 1096 AGTCCGCAACAGCTCCAGCAATCAATTTGAT-----GGGAAATGGTCAAGAGCT 1149
 QY 380 ValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPhe 399
 Db 1150 GTTCGAAAGATGAGCATGGTATGCTTTTGAAGAGATGAGTTCCTGTTATATCCCA 1209
 QY 400 AlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGlu 419
 Db 1210 GCCAAGATCTTTCAGCAGAAACAGCAGAGGCGATTTGATAGCAATGCGCCAGCAGGAA 1269
 QY 420 SerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPhe 439

Db 1270 AGTTTATCTCATAACTAGGAACTAAGAAACTGACCTCCCATCTAGTCATGAGAAATTT 1329
 QY 440 TyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGly 459
 Db 1330 TACAATAGGCTTATGACTTACTAGCAAGATTACCAAGATTTTACTTGTATATTAAGGT 1389
 QY 460 ArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThr 479
 Db 1390 CGACAAAGTTGATTTTGGAGCTTTGGATAACCTGTTGGAAAGCACTCAAGGATGCTCAAGT 1449
 QY 480 AsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGlu 499
 Db 1450 GATAAAGTCAAGTTAGTGGAGATATCTTGCCTTCTTAGCTCCGATTCGTCATCCAGAA 1509
 QY 500 ArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGln 519
 Db 1510 CGTTTAGGAAACCAATGCGCAATTAACCTACACTGATGATGAGATTCAAGTAGCCCAAG 1569
 QY 520 LeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSer 539
 Db 1570 TTGGCAGGCAAGTACACAGCAGAGAGCGTTATATCTTTGATCCTCGTGATATACCACT 1629
 QY 540 AspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAsp 559
 Db 1630 GATGAGGGGGATGCTATGTAACTCCACATATGACCCATGACCCACTGGATTAAAAAAGAT 1689
 QY 560 SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeu 579
 Db 1690 AGTTTGTCTGAGCTGAGAGCGCGACCCGAGCTTATGCTRAAGAGAAAGGTTTGACC 1749
 QY 580 ProProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIle 599
 Db 1750 CCTCCTTCGACAGACCATCAGGATTCAGGAAATACTGAGGCAAAAGGAGCAGAGACTATC 1809
 QY 600 TyrAsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGlu 619
 Db 1810 TACAACCGMTGAAAGCAGCTTAAGAGGTGCACTTGTATGCTGCTTACAATCTTCAA 1869
 QY 620 HisThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsn 639
 Db 1870 TATACGTAGAGTCAAAAACGGTAGTTTATCATACCTCATTTATGACCATTTACCATTAAC 1929
 QY 640 IleLysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGlu 659
 Db 1930 ATCAAAATTTGAGTGGTTTGACGAAGGCTTTATGAGGCACCTAAGGGGTATATCTTTGAG 1989
 QY 660 AspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsn 679
 Db 1990 GATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAACGACGTCGCATTCAGAT 2049
 QY 680 AspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspPro 699
 Db 2050 AATGGTTTGGTAACGCTAGCGACCAATGTTCAAGAAACAAATAATGGTCAAGCTGATACC 2109
 QY 700 AsnLysAsnPheLysAlaAspGluGlu-----ProValGluGluThr 713
 Db 2110 AATCAACCGGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACCTGAGGAGAAACC 2169
 QY 713 ----- 713
 Db 2170 CCTCGAAGAGAAACCCGAAAGCGAGAAACAGAGTCTCCAAAAACCAACAGAGGAACCA 2229
 QY 714 -----ProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGln 730
 Db 2230 GAAGAAATCACAGAGGATCAGAAAGAACCTCAGTCGACACTGAAAGAGTTGAAGAAAAA 2289
 QY 731 LeuLysGluAlaGluValLeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAla 750
 Db 2290 CTGAGAGAGGCTGAAGATTTACTTGGAAAAATCCAGGATCCAAATATCAAGTCCATGCC 2349
 QY 751 ThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSer 770

Db 2350 AAGAGAGCTCTCAGAGTAATAAAAAATAATTACTATTGGCACCAGGACCAACAATACT 2409

Qy 771 ILeMetAlaGluAlaGluLysLeuLeuAlaLeuLeuLysGlySer 785

Db 2410 ATTATGGCAGAGCTGAAGAACTATTGGCTTTATTAAGAGAGAGT 2454

RESULT 11

US-10-387-783-11

; Sequence 11, Application US/10387783

; Publication No. US20040005331A1

; GENERAL INFORMATION:

; APPLICANT: Johnson, Leslie E.

; APPLICANT: Adamou, John E.

; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus

; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural

; TITLE OF INVENTION: Motifs

; FILE REFERENCE: 469201-683

; CURRENT APPLICATION NUMBER: US/10/387,783

; CURRENT FILING DATE: 2003-03-13

; PRIOR APPLICATION NUMBER: 09/468,656

; PRIOR FILING DATE: 1999-12-21

; PRIOR APPLICATION NUMBER: 60/113,048

; PRIOR FILING DATE: 1998-12-21

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 3.0

; SEQ ID NO 11

; LENGTH: 2531

; TYPE: DNA

; ORGANISM: Streptococcus pneumoniae

US-10-387-783-11

Alignment Scores:

Pred. No.:	4,01e-243	Length:	2531
Score:	2786.50	Matches:	544
Percent Similarity:	77.06%	Conservative:	84
Best Local Similarity:	66.75%	Mismatches:	140
Query Match:	66.90%	Indels:	47
DB:	16	Gaps:	7

US-09-765-271-56 (1-796) x US-10-387-783-11 (1-2531)

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnAsnArgValSer 19

Db 61 TCTATGAGCTGGAGCTTACCAAGCTGGTCAGGATAAGAAAGAGCTTAATCGAGTTGCT 120

Qy 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39

Db 121 TATATAGATGGTCATCAGGCTGGTCAAAAGGCGCAAAACTTGACACGATGAAGTCAGT 180

Qy 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59

Db 181 AAGAGGGAGGGGATCAACCGCAACAAATTTGTTATCAAGATTACGATCAGGTTATGTG 240

Qy 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIle 79

Db 241 ACCTTCATCGAGACCATATCATCTACTATAATGCGAAGTTCCTTATGATGCCATCATC 300

Qy 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99

Db 301 AGTGAAGAGCTCCTCATGAAGATCCGAATATATCAGTTGAAGATTACAGACATGTGCAAT 360

Qy 100 GluValLysGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119

Db 361 GAAATCAAGGGTGGTATGTCATTAAAGTAAACCGTAAATACTATGTATTACCTTAAGAT 420

Qy 120 AlaAlaHisAlaAspAsnValArgThrLysGluLysGluLeuAsnArgGlnLysGlnGluHis 139

Db 421 GCRCTCATCGGATATATTCGGACAAAAGAGAGATTAACCGTCAGAAGCAGAAACGC 480

Qy 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159

Db 481 AGTCATATCATATACTCA-----AGACGAGATAATGCTGTGTGCGCCAGAGCC 531

Qy 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179

Db 532 CAAGGAGCTTATATCAACCGATGATGGGTATATCTTCAATGCTATGATATCATTTGAGGAC 591

Qy 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199

Db 592 ACGGCTGATGCTTATATCGTTCTCAGCGGACCATTTACCATTTACATTTCTTAAGATGAG 651

Qy 200 LeuSerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219

Db 652 TTATCAGCTAGGAGTAGCTGCTGAGAACCTATTATGGAATGG----- 696

Qy 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239

Db 697 -----AAGCAGGATCTCGTCTCTTCAAGTCTAGTTATATAATGCA 738

Qy 240 SerValSerAsnProGlyThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259

Db 739 AATCAGCTCAACCAAGATTGTCAGAGAACACATCTGACTGTCACCTCAACTTATCAT 798

Qy 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279

Db 799 CAA---AATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTGTATGCTAAACCTTA 855

Qy 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299

Db 856 TCAGAACGCCATGTGGAATCTGATGGCCTTATTTTCGACCCAGCCCAATCAAGTCA 915

Qy 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319

Db 916 ACCGCCAGAGGTGTAGCTGCTCCTCATGTGTAACATTACCATTTATCCCTTATGAACAA 975

Qy 320 MetSerGluLeuGluGluAlaArgIleLeuProLeuArgTyrArgSerAsnHis 339

Db 976 ATGCTCTGAATGGAAAAAGATTGCTCGTATTAATTCCTTCGTTATCGTTCAACCAAT 1035

Qy 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359

Db 1036 TGGGTACCATGATTCAAGACCAAGAACCAAGTCCACACCCAGCTCCAGACCTAGTCCA 1095

Qy 360 GlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeu 379

Db 1096 AGTCCGCAACCAAGCTTCAAGCAATCCAATTGAT-----GGGAAATTTGTCAAAGAGCT 1149

Qy 380 ValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPhe 399

Db 1150 GTTCGAAAAGTAGGCGATGTTATGCTTTTGAAGAGATGGAGTTTCTCGTTATATCCCA 1209

Qy 400 AlaLysAspLeuProSerGluThrValLysAsnLeuLysSerLysLeuSerLysGlnGlu 419

Db 1210 GCCAAGGATCTTTCAGCAAAAACAGCAGCAGGCGATTGTAGCAAACTGGCCACAGCAGAA 1269

Qy 420 SerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPhe 439

Db 1270 AGTTTATCATTAAGCTAGGAATAAGAACTAGAACCTCCCATCTAGTGTATGATGAGATTT 1329

Qy 440 TyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe---AsnLysGly 459

Db 1330 TACAATAAGCTTATGACTTACTAGCAAGAAATTCACCAAGATTACTTGTATAATAAGGT 1389

Qy 460 ArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThr 479

Db 1390 CGACAAAGTTGATTTTGGAGTTTGGATTAACCTGTGGAACGATCAAGGATGCTCAAGT 1449

Qy 480 AsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGlu 499

Db 1450 GATAAAGTCAAGTTAGTGGAGATATTCTTTCCTTCTTAGCTCCGATTCGTCATCCAGAA 1509

Qy 500 ArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValAlaGlnIleAlaGln 519

Db 1510 CGTTTAGAAAACCAAAATGGCAAAATTACCTACATGATGATGAGATTCAAGTAGCCAG 1569

Qy 520 LeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSer 539


```

Db 1570 TTGGCAGGCAAGTACACAGCAGAGCGGTATATCTTTGATCCTCGTGATATACCACT 1629
Qy 540 AspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTripleGlyLysAsp 559
Db 1630 GATGAGGGGATGCTCTATGTAACTCCATATGACCCATAGCCACTGGAATAAAAAAGAT 1689
Qy 560 SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyLysLeu 579
Db 1690 AGTTTGCTGAGCTGAGAGCGCAGCCCGAGGCTTATGCTTAAGAGAGAAAGTTTGACC 1749
Qy 580 ProProSerProAspAlaAspValLysAlaAlaProThrGlyAspSerAlaAlaAlaIle 599
Db 1750 CCTCTTCGACAGCATTACAGGATTCAGAAATATCTGAGCAAAAGAGGAGCAAGCTATC 1809
Qy 600 TyrAsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGlu 619
Db 1810 TACAAACCGGTGAAGAGCAGCTAAGAAAGGTCACCTGATGCTGATGCTTACAAATCTCAA 1869
Qy 620 HisThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsn 639
Db 1870 TATATCTGTAAGTCAAAACCGTAGTTTATCATCTCTATTATGACCATTTACCATTAAC 1929
Qy 640 IleLysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGlu 659
Db 1930 ATCAAAATTTGAGTGGTTTGACGAAGGCTTTTATGAGGCACCTAAAGGGGTATACTCTTGAG 1989
Qy 660 AspLeuPheAlaTyrIleLysTyrValGluHisProAspGluArgProHisSerAsn 679
Db 1990 GATCTTTGGCACTGTCAAGTACTATGTGCAATCCAAACGACGCTCCGCATTCAGAT 2049
Qy 680 AspGlyTrpGluAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspPro 699
Db 2050 AATGGTTTGGTAACGCTAGCGACCATGTTCAAGAAACAAATGTTGTCAGCTGATACC 2109
Qy 700 AsnLysAsnPheLysAlaAspGluGlu-----ProValGluGluThr 713
Db 2110 AATCAACGGAAACCAACGAGCGAGGAGAAACCTCAGACAGAAACCTGAGGAGAAACC 2169
Qy 713 ----- 713
Db 2170 CCTCGAAGAGAAACCGAAGCGAGAAACCAAGCTCTCCAAACCAACAGAGGAAACCA 2229
Qy 714 -----ProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGln 730
Db 2230 GAAGAAATCACAGAGGAATCAGAGAACCTCAGGTCGAGACTGAAAGGTTGAAGAAAA 2289
Qy 731 LeuLysGluAlaGluValLeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAla 750
Db 2290 CTGAGAGGCTGAAGATTTTACTTGGAAAAATCCAGGATCCAAATTAAGTCCAATGCC 2349
Qy 751 ThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSer 770
Db 2350 AAAGAGACTCTCAGAGATTAAAAATTAATTTACTATTGGCCACCCAGGACAAACAATACT 2409
Qy 771 IleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer 785
Db 2410 ATTATGGCAGAGCTGAAGAACTATTGGCTTTATTAAGAGAGAGT 2454

```

RESULT 12

```

US-10-412-862-5
; Sequence 5, Application US/10412862
; Publication No. US20040052781A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; FILE REFERENCE: 469201-685
; CURRENT APPLICATION NUMBER: US/10/412,862
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656

```

```

; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-412-862-5

```

```

Alignment Scores:
Pred. No.: 8,34e-242 Length: 2531
Score: 2772.00 Matches: 541
Percent Similarity: 75.78% Conservative: 91
Best Local Similarity: 64.87% Mismatches: 142
Query Match: 66.55% Indels: 60
DB: 13 Gaps: 8

```

US-09-765-271-56 (1-796) x US-10-412-862-5 (1-2531)

```

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19
Db 61 TCCTATGAAGTCTGCTCGTCCACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTTCT 120
Qy 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
Db 121 TATATAGATGCTGATCAGGCTGGTCAAAAGGCAGAAAACTTGACACAGATGAAGTCAGT 180
Qy 40 LysArgGluGlyLysAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
Db 181 AAGAGGAGGAGGAGTCAACGCGCAACAAATCGTCAATCAAGATTACGATCAAGGTTATGTG 240
Qy 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIle 79
Db 241 ACCTCTCATGAGACCATTAATCATTAATTAAGCAAGGTCCTTTATGATGCCATCATC 300
Qy 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
Db 301 AGTGAAGAGCTCCTCATGAAAGATCCGAATTATCAGTTGAAGGATTTCAGACATTTGCAAT 360
Qy 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
Db 361 GAAATCAAGGCTGTTATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGAT 420
Qy 120 AlaAlaHisAlaAspAsnValArgThrLysGluGlnIleAsnArgGlnLysGlnGluHis 139
Db 421 GCAGCTCATCGGATTAATTCGACAAAGAAAGAGATTAAACGTGAGAGGAGGAAACAC 480
Qy 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyValAlaLeuAlaArgSer 159
Db 481 AGTCATAAATCACGGGGTGGTTCT-----AACGATCAAGCAGTAGTTGCGAGCCAGGCC 534
Qy 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
Db 535 CAAAGGACGCTATACACGAGTATGTTATCTTCAATGCAATCTGATATCATTTAGGAC 594
Qy 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
Db 595 ACGGTTGATGCTTATATCTTCTCAGCGGACCATTAATTAATTAATTAATTAATTAATTAAT 654
Qy 200 LeuSerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
Db 655 TTATCAGCTAGCGAGTTAGTCTGTCAGAAAGCTTATGGAATGGG----- 699
Qy 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239
Db 700 -----AAGCAGGAGTCTCGCTCTCTTCTTCAAGTTCTAGTTATAATGCA 741
Qy 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
Db 742 AATCCAGCTCAACCAAGATTGTCAGAAACCAAACTCTGACTGTCTCACTCAACTTATCAT 801

```

260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
 802 CAA---AATCAAGGGGAAACATTTCAAGCGCTTTTGAATTTGTATGCTTAACCCCTTA 858
 280 SerClnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
 859 TCAGAACGCCAATGGAATCTGATGGCTTATTTTCGACCCAGCGCAATCACAAAGTGA 918
 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
 919 ACCGCCAGAGGTGATGCTGCTCATGGTAACCATTAACCACTTTATCCCTTATCAACAA 978
 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
 979 ATGTCGTAATTTGGAACAAAGCAATTTCTCGTATTTATCCCTTCGTTATCGTTCAACCAT 1038
 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
 1039 TGGGTACCAAGATTCAAGACCAAGAACCAACCAAGTCCCAATCGACTCCGGAACCTTAGTCCA 1098
 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373
 1099 AGTCGCAACCTGACCAACAAATCTCAACAGCTCCCAAGCAATCCCAATTCAT 1152
 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393
 1153 AAATTTGGTCAAGAAAGCTGTTTCGAAAGTAGGCGATGTTATGCTTTTGGAGGAATGGA 1212
 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413
 1213 GTTCTCGTTATATCCAGCCAGAGATCTTTCAGCAGAAACAGCAGCAGGCAATTCATAGC 1272
 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433
 1273 AAATGGCCACAGCAGAAAGTTATCTCATAGCTAGGAGCTTAAGAAACTGACCTCCCA 1332
 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453
 1333 TCTAGTGTACGAGATTTTACAAATAGGCTTATGACTTACTAGCAAGAAATTCACCAAGAT 1392
 454 LeuPhe**AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473
 1393 TTACTTGATATAAAGTTCGACAGATTTGATTTTGGAGCTTTGGATAACCTGTTGGAACGA 1452
 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla 493
 1453 CTCAGGATGTCCTCAAGTATGATAAAGTCAAGTTAGTGGATGATATCTTCCTCTTAGCT 1512
 494 ProlThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513
 1513 CCGATTGCTATCCAGAACGTTTAGGAAACCAATGCGCAATTAACCTACACTGATGAT 1572
 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533
 1573 GAGATTCAAGTACCAAGTTGGCAGGCAAGTACACAAACAGAACGCGTTATATCTTTGAT 1632
 534 GluHisAspIleLeuSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553
 1633 CCTCGTATATAACCAAGTATGAGGGGGATGCTATGTAACCTCCATATGACCCATAGC 1692
 554 HisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr 573
 1693 CACTGATTAATAAAGATAGTTTGTCTGAAGCTGAGAGCGCGCAGCCGCTTATGCT 1752
 574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593
 1753 AAAGAGAAAGGTTTGCACCCCTCTTCACAGACCATCAGGATTTCAGGAATACTAGGCA 1812
 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArg 613
 1813 AAAGGAGCAAGAGCTATCTCAACCCGCTGAAGACAGCTAAGAGAGGTGCCACTTATCGT 1872
 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuLeuIleProHis 633

1873 ATGCTTACAAATCTTCAATATACTGTAGAAGTCAAAACCGTAGTTTAAATCATACCTCAT 1932
 634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653
 1933 TATGACCATTAACCAATCAATATTTAGTGGTTTACGAAGCCCTTATGAGGCACCT 1992
 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673
 1993 AAGGGGTATCTCTTGGAGATCTTTTGGCGACTCTCAAGTACTATGTGCAACATCCAAAC 2052
 674 GluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLys--- 692
 2053 GAACGTCGGATTCAGATAATGTTTGGTAACCTAGCGACCATGTTCGTAATAAATAAG 2112
 693 ---LysAspHisSerGlu--- 697
 2113 GTAGACCAAGACAGTAACCTGATGAAGATAAGAAACATGATGAAGTAAGTGAGCCAACT 2172
 697 --- 697
 2173 CACCTGAACTGATGAAAAAGAGAAATCACGCTGGTTTAAATCTTTCAGCAGATAATCTT 2232
 698 ---AspProAsnLysAsnPheLysAlaAspGluGluProValGluThrProAlaGlu 716
 2233 TATAAACCAAGCACTGATACGGAAGAGACAGAGAGAAAGCTGAAGATACACACATGAG 2292
 717 ProGluValProGlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluVal 736
 2293 GCTGAAATCTCTCAAGTAGAAGAAATCTCTTATTAACGCTAAGATAGCAGATGCGGAGCC 2352
 737 LeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGly 756
 2353 TTGCTAGAAAAAGTAACAGATCTCTAGTATTAGACAAAAATGCTATGAGACATTTGACTGT 2412
 757 LeuArgAsnAsnLeuThrIleGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGlu 776
 2413 CTAAAAAGTAGTCTCTTCTCGAAACGAAAGATAATAACACTATTTTCAGCAGATAGAT 2472
 777 LysLeuLeuAlaLeuLeuLysGlySerAsnProSerSerVal 790
 2473 AGTCTCTTGGCTTTGTTAAAGAAAGTCAACCGGCTCTCTATA 2514

RESULT 13

US-10-412-850-5
 ; Sequence 5, Application US/10412850
 ; Publication No. US20040001836A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-686
 ; CURRENT APPLICATION NUMBER: US/10/412,850
 ; CURRENT FILING DATE: 2003-04-14
 ; PRIOR APPLICATION NUMBER: 09/468,656
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: Patent In Ver. 3.0
 ; SEQ ID NO 5
 ; LENGTH: 2531
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-10-412-850-5

Alignment Scores:
 Pred. No.: 8,34e-242 Length: 2531
 Score: 2772.00 Matches: 541
 Percent Similarity: 75.78% Conservative: 91
 Best Local Similarity: 64.87% Mismatches: 142

Query Match:	66.55%	Indels:	60
DB:	16	Gaps:	8
US-09-765-271-56 (1-796) x US-10-412-850-5 (1-2531)			
QY	1	SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---	LysGluAsnAsnArgValSer 19
DB	61	TCCTATGAACCTTGGTTCGTCACCAAGCTGGTCAGGTTTAAGAAAGAGCTCTAATCGAGTTCT	120
QY	20	TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer	39
DB	121	TATATAGATGGTATCAGCGTGGTCAAAAGGACAGAAACTTGCACACCAAGATCAAGTCAGT	180
QY	40	LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal	59
DB	181	AAGAGGAGGGGATCAACGCCGCAACAAATCGTCATCAAGATTACGGATCAAGGTTATGTG	240
QY	60	ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle	79
DB	241	ACCTCTCATGGAGACCAATTATCATTAATAATGCAAGGTCCCTTATGATGCCATCATC	300
QY	80	SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn	99
DB	301	AGTGAAGAGCTCCTCATGAAGATCCGAAATTATCAGTTGAAGSANTCAGACATTTGTCAAT	360
QY	100	GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp	119
DB	361	GAATCAAGGGTGGTTATGTTATCAAGGTAGTCAAGAAATACTATGTTTACCTTAAGGAT	420
QY	120	AlaAlaHisAlaAspAsnValArgThrLysGluIleAsnArgGlnLysGlnGluHis	139
DB	421	GCAGCTCATCGGATATAATTCGGACAAAAGAGAGATTAAACCTCAGAAGCAGAAACAC	480
QY	140	SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer	159
DB	481	AGTCATAATCACGGGGTGGTTCT-----AACGATCAAGCAGTAGTTGCAGCCAGAGCC	534
QY	160	GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp	179
DB	535	CAAGGACGCTATACACGGATGATGGTTATCTTCATCGACTGATATCATTTGAGGAC	594
QY	180	ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu	199
DB	595	ACGGGTGATGCTTATATCGTTCCTCAACGGCCACCATTACATTACATTCCTAAGATGAG	654
QY	200	LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer	219
DB	655	TTATCAGCTAGCCAGTTAGTCTGCTCAGAAAGCCTATTGGAATGGG-----	699
QY	220	AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro	239
DB	700	-----AAGCAGGGATCTCGTCCTCTCTCAAGTTCTAGTTATAATGCA	741
QY	240	SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer	259
DB	742	AATCCAGCTCAACCAAGATTGTGCAGAAACCAATCTGACTGTCTCACTCCAACTTATCAT	801
QY	260	GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu	279
DB	802	CAA---AATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTTGATGCTAAACCCCTTA	858
QY	280	SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg	299
DB	859	TCAGAACGCCCATGTGGAATCTGATGGCTTATTTTCGACCCAGCGCAATCAAGTCGA	918
QY	300	ThrAlaArgGlyValAlaValProHisTyrHisTyrHisPheIleProTyrSerGln	319
DB	919	ACCGCAGAGGTGATGCTGCTCCCTCATGTGTAACCATTAACCTTTATCCCTTATGACAA	978
QY	320	MetSerGluLeuGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis	339
DB	979	ATGTCCTGAATTCGAAACGAATTCGCTATATTATCCCTTCGTTATGCTTCAAAACCAT	1038
QY	340	TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro	359
DB	1039	TGGGTACCAGATTCAAGACCAAGAACCAAGTCACAAATCGACTCCGGAACCTAGTCCA	1098
QY	360	GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer	373
DB	1099	AGTCGCAACCTGCACCAAAATCCTCAACAGCTCCAAAGAAATCCAAATTTGAT-----GAG	1152
QY	374	SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly	393
DB	1153	AAATTGGTCAAGAGAGCTGTTTCAAAAGTAGCGATGGTTATGTTCTTTTGGAGGAATCGA	1212
QY	394	IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer	413
DB	1213	GTTTCTCGTTATATCCAGCCCAAGGATCTTTTCAGCAGAAACAGCAGCAGGCTTATGATAGC	1272
QY	414	LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla	433
DB	1273	AAACTGCCAAGCAGGAAGTTTATCTCAATAAGCTTAGGAGCTAAGAAACTGACCTCCA	1332
QY	434	ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla	453
DB	1333	TCTAGTCATCGAGAAATTTTACAATAAGCTTATGACTTACTAGCAAGAAATTCACCAAGAT	1392
QY	454	LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg	473
DB	1393	TTACTTGATAATAAAGGTCGACAAAGTTGATTTTGAGGCTTTGGATAACCTGTTGGAACGA	1452
QY	474	LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla	493
DB	1453	CTCAAGGATGTCCTCAAGTGAATAAGTCAAGTTAGTGGATGATATTTCTTGGCTTCTTAGCT	1512
QY	494	ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp	513
DB	1513	CCGATTCTCATCCAGAACGTTTAGGAAACCAATTCGCCAAATTACCTACACTGATGAT	1572
QY	514	GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp	533
DB	1573	GAGATTCAAGTACCAAGTTGGCAGGCAAGTACACAAACAGACGCTTATATCTTTGAT	1632
QY	534	GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer	553
DB	1633	CCTCGTGAATATACCAAGTATGAGGGGATGCTCTATGTAACTCCACATATGACCCATAGC	1692
QY	554	HisTyrIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr	573
DB	1693	CACGCTGATTAAAAAAGATAGTTTCTGAAAGCTGAGAGCGGCGCCAGGCTTATGCT	1752
QY	574	LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly	593
DB	1753	AAAGAGAAAGGTTTGACCCCTCTTCAGACAGACCATCAGGATTCAGGAAATCTAGAGCA	1812
QY	594	AspSerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArg	613
DB	1813	AAAGAGCAGAGACTATCTACAAACCGCGAAGCAGCTAAGAAAGGCGCCATGATCGT	1872
QY	614	LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis	633
DB	1873	ATGCTCTTACAATCTTCAATATATCTAGAGAGTCAAAACCGTAGTTTATATCATCTCAT	1932
QY	634	LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro	653
DB	1933	TATGACCATTAACATAACATCAAAATTTGAGTGGTTTTCAGAAAGGCTTTTATGAGCACCT	1992
QY	654	AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp	673
DB	1993	AAGGGGTATCTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCAAAAC	2052
QY	674	GluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLys---	692
DB	2053	GAACGTCGCAATTCAGATAATGGTTTTCGTAACGCTAGCAGCACTGTTTCGTAATAAATAAG	2112
QY	693	-----LysAspHisSerGlu-----	697

60 ThrSerHisGlyAspHisTyrHisTyrTrpAsnGlyLysValProTyrAspAlaIlelle 79
241 ACCTCTCATGGAGACCAATATCATTTACTATATAATGCGAAGGTCCCTTATGATGCCATCATC 300

80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValasn 99
301 AGTGAAGAGCTCTCATGAAGATCCGAATTTATCATGTTGAAGGATTCAGACATTTGTTCAAT 360

100 GluValLysGlyGlyTyrValLleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
361 GAAATCAAGGGTGGTTATGTTATCATCAAGGTAGATGGAATACTATGTTTACCTTAAAGGAT 420

120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
421 GCAGCTCATGCCGATATATTCGRCAAAAGAAGAGATTAAACGTCAGACGAGGAACAC 480

140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
481 AGTCATAAATCACGGGGGTGGTTCT-----AACGATCAAGCATGATGTGACCGACAGAGCC 534

160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
535 CAAGGACGCTATACAACGAGATGATGTTATATCTTCATGTCATCTGATATCATTTGAGGAC 594

180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
595 ACGGTGATGCTTATATCGTTCCTCAGCGGACCATTACATTTACATTCCTAAGAATGAG 654

200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
655 TTATCAGCTAGCAGTTAGTCTGTCAGAGGCTATTGGAATGGG----- 699

220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239
700 -----AAGCAGGGATCTCGTCTCTTCAAGTTCATGTTATAATGCA 741

240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
742 AATCCAGCTCAACCAGATTTGTCAGAGAACCAACAATCTGACTGCTCACTCCAATTTATCAT 801

260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
802 CAA---AATCAAGGGGAAAACATTTCAAGCCTTTTACGTGAATGTATGCTAAACCCCTTA 858

280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
859 TCAGAACGCCATGTGGAATCTGATGGCTTATTTTCGACCGAGCGAAATFACAAGTCGA 918

300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
919 ACCGCGACAGGTGTAGCTGTCCCTCATGTGAACCAATACCACCTTATCCCTTTATGACAA 978

320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
979 AGTCTGHAATTTGGAAAAACGAATGCTCGTATTTATTCCTCTGTTATCGTTCAAAACCAT 1038

340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
1039 TGGGTACCAGATTCAGACACAGAACCAACAGTCCACAATCGACTCCGGAACCTTAGTCCA 1098

360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373
1099 AGTCGCGCAACCTGCGACCAAAATCTCAACGACTCAAGCATCAATGAT-----GAG 1152

374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393
1153 AAATTTGGTCAAGAAGCTGTTTCAAAAAGTAGCGCATGTTATGTTCTTTGAGGAGAAATGGA 1212

394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413
1213 GTTTTCGTATATCCAGCCCAAGGATCTTTCAGCGAAGAACAGCAGCGCATTTGATGAC 1272

```

RESULT 14
US-10-387-783-5
; Sequence 5, Application US/10387783
; Publication No. US2004000531A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural

```

200	LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerClyArgGlyAsnLeuSer	219
655	TTATCAGCTACGAGGTAGTCTGTCAGAAAGCCTATTGGAATGGG-----	699
220	AsnSerArgThrTyArgAGGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro	239
700	-----AAGCAGGGATCTCGTCCCTCTTCAAGTCTAGTTATAATGCA	741
240	SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer	259
742	AATCCAGCTCAACCACGATTGTCAGAAACCAACAATCTGACTCTCACTCCAACCTTATCAT	801
260	GlnAlaSerGlnSerAsnAspIleAaspSerLeuLeuLeuGlnLeuTyryLysLeuProLeu	279
802	CAA---AATCAAGGGGAAAAACATTCAAGGCTTTTACGTGAATGTATGCTAAACCCCTTA	858
280	SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg	299
859	TCAGAACGCCATGTGGAACTGTGATGGCTTATTTCGACCCAGCGAAATCACAAGTCGA	918
300	ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln	319
919	ACCGCCAGAGGTGTAGTGTGCTCCCTCATGTGAACCATTACCACCTTATCCCTTATGAGCAAA	978
320	MetSerGluLeuGluGluAargIleAlaArgIleIleProLeuArgTyrArgSerAsnHis	339
979	ATGTCGTGAATTGGAAAAACAAAGTAATGTCGTATTATTCCTCTCGTTATCGTTCAAACCAT	1038
340	TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro	359
1039	TGGGTACCAAGATTCAAGACCAAGAACCAACCAAGTCCCAATCGACTCCGGAACCTAGTCCA	1098
360	GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer	373
1099	AGTCGCGCAACCTGCACCAAAATCCTCAACCAAGCTCAAGCAATCCAATGAT-----	1152
374	SerLeuValSerGlnLeuValArgLysValGlyGluTyrValPheGluGluLysGly	393
1153	AAATTGGTCAAGAAGCTGTTCGAAAAAGTAGCGATGTTATGTTCTTTGAGGAGAAATGGA	1212
394	IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer	413
1213	GTITTCGTATTATCCCAAGCATCTTTTCAGCGAAAAACAGCAGCAGGCAAGCATTTGATGC	1272

QY	320	MetSerGluLeuGluGluArgGileAlaAArgIlelleProLeuArgTyrArgSerAsnHis	339
Db	979	ATGTCGAATTGGAAAAACGAATTGCTCGTATATTATCCCTTCGTATCGTTCAAAACCAT	1038
QY	340	TrpValProAspSerArgProGluGluProSerProGlnProThrProGluProSerPro	359
Db	1039	TGGGTACCAGATTCAAGACCAAGACCAACCAAGTCCCAATCGACTCCGGAACCTAGTCCA	1098
QY	360	GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer	373
Db	1099	AGTCGCGCAACTGCACCAAAATCTCAACGAGCTCCAAGCAATCAAAATTGAT-----GAG	1152
QY	374	SerLeuValSerGlnLeuValAlaArgLysValGlyGluGlyTyrValPheGluGluLysGly	393
Db	1153	AAATTGGTCAAAAGAGCTGTTCCGAAAGTAGCGATGGTTATGCTTTGAGGACAATGGA	1212
QY	394	IleSerArgTyrValPheAlaIleAspLeuProSerGluThrValLysAsnLeuGluSer	413
Db	1213	GTTTCTCGTTATATCCAGCCAAAGATCTTTTCAGCAGAAACAGCAGCAGGCAATTGATAGC	1272

Qy	360	GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer	373
Db	1099	AGTCGCGAACCTGCGACCAATCCTCAACCAAGCTCCAAGCAATGAT-----GAG	1152
Qy	374	SerLeuValSerLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly	393
Db	1153	AAATTGGTCAAGAAGCTGTTCCGAAAGTAGGCGATGTTATGTTCTTTGAGGAGAATGGA	1212
Qy	394	IleSerArgTyrValPheAlaIleAspLeuProSerGluThrValLysAsnLeuGluSer	413
Db	1213	GTTTCTCGTTATATCCCGCAAGGATCTTTTCAGCGAAGAACAGCAGCGAGGCATTTGNTAGC	1272

Db	1099	AGTCCGCAACCTGCACCAAAATCCTCAACGAGCTCCAAGCAATCCAATTGAT-----GAG	1152
QY	374	SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly	393
Db	1153	AAATTGGTCAAGAAGCTGTTCCGAAAGTAGCGCATGGTTATGCTTTGAGGAGAAATGGA	1212
QY	394	IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuLysSer	413
Db	1213	GTTTTCGTTATATCCCGCAAGGATCTTTACGAGAAACAGCAGCGAGCGATTGATAGC	1272

QY 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysLysGluAsnValAla 433
Db 1273 AAACCTGGCCAGCAGAAAGTTTATCTCATAGCTAGGAGCTTAAGAAACTGACCTCCA 1332
QY 434 ProArgGlnGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453
Db 1333 TCTAGTGTGCGAAGATTTTACAAATAGGCTTATGACTTACTAGCAAGAAATTCACCAAGAT 1392
QY 454 LeuPhe***AsnLysGlyArgAsnSerPheGlnAlaLeuAspLysLysLeuLeuLys 473
Db 1393 TTACTTTGATATAAGAGTCGCAAGTTGATTTTGGAGCTTTGGATAACCTGTTGAAACGA 1452
QY 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspLeuLeuAlaPheLeuAla 493
Db 1453 CTCAGGATGTCCCAAGTGAATAAGTCAAGTTAGTGGATGATATCTTTCCTTCTAGCT 1512
QY 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513
Db 1513 CCGATTCTGTCATCCAGAACGTTTAGGAAACCAATGCGCAATTTACCTACACTGATGAT 1572
QY 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533
Db 1573 GAGATTCAGAGTAGCAAGTTGCGAGCGAAGTACACACAGAGAGCGGTATATCTTTGAT 1632
QY 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553
Db 1633 CCTCGTGATATAACAGTGTAGGAGGATGCTTATGTAATCTCCACATATGACCATAGC 1692
QY 554 HistripIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr 573
Db 1693 CACTGGATTAAAGAGTAGTTTCTGGAAGCTGAGAGCGGAGCGCCAGGCTTATGCT 1752
QY 574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593
Db 1753 AAAGAGAAAGTTTGACCCCTCTTCACAGACCATCAGGATTCAGGAAATACCTGAGCCA 1812
QY 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg 613
Db 1813 AAAGGAGCAGAAAGCTATCTCAACCGCGTGAAAGCAGCTAAGAGGTCACCTTGATCGT 1872
QY 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633
Db 1873 ATGCTTTACAACTCTTCAATATCTGTAAGAGTCAAAACCGGTAGTTTAAATCACTCAT 1932
QY 634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653
Db 1933 TATGACCATTACCATAACATCAATTTGAGTGTTCAGAGCCCTTTATGAGGACCT 1992
QY 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673
Db 1993 AAGGGGTATCTCTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAAC 2052
QY 674 GluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLys--- 692
Db 2053 GAACGTCGGCATTCAGATTAATGTTTGGTAAACGCTAGCAGACCATGTTTCGTAATAAATAG 2112
QY 693 -----LysAspHisSerGlu----- 697
Db 2113 GTAGACCAACAGACAGTAACTGTGATGAAGTAAAGAACATGATGAAGTGAAGCAACT 2172
QY 697 ----- 697
Db 2173 CACCTCGAATCTGATGAAAGAAAGAAATCACGCTGGTTTAAATCCTTCAGCAGATATCTT 2232
QY 698 ---AspProAsnLysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGlu 716
Db 2233 TATAACCAAGCAGCTGATCGGAAGAGACAGAGAAAGAGCTGAAGATACCAAGATGAG 2292
QY 717 ProGluValProGlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluVal 736
Db 2293 GCTGAAATTCCTCAAGTACAGAAATCTGTTTAAACGCTTAAGATACAGATCGAGCGGCC 2352
QY 737 LeuLeuAlaLysValThrAspSerSerLysLysAlaAsnAlaThrGluThrLeuAlaGly 756

Db 2353 TTGCTAGAAAAGTAACAGATCCTAGTATTAGACAAAATGCTATGAGACATTTACTGCT 2412
QY 757 LeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGlu 776
Db 2413 CTAAAAGTAGTCTTCTCTCGAAGCAAGAAATATAACACTATTTTACAGAAAGTAGAT 2472
QY 777 LysLeuLeuAlaLeuLeuLysGlySerAsnProSerVal 790
Db 2473 AGTCTCTTGGCTTTGTTAAAGAAAGTCAACCGGCTCTCTATA 2514

RESULT 15
US-09-884-465A-5
; Sequence 5, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 2639
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-5

Alignment Scores:
Pred. No.: 7,96e-240 Length: 2639
Score: 2750.50 Matches: 538
Percent Similarity: 75.30% Conservative: 90
Best Local Similarity: 64.51% Mismatches: 145
Query Match: 66.04% Indels: 61
DB: 10 Gaps: 8

US-09-765-271-56 (1-796) x US-09-884-465A-5 (1-2639)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19
Db 174 TCCTATGAACTTGGTCTGTCACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTCT 233
QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
Db 234 TATATAGATGGTGATCAGGCTGGTCAAAAGGCGAGAAAATTTGACACAGATGAAGTCAGT 293
QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
Db 294 AAGAGAGAGGGGATCAACCGCCGAACAAATTTGTTATCAAGATTACGGATCAAGGTTATGTG 353
QY 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIle 79
Db 354 ACCTCTCATGGAGACCATTTATCTACTATATATGCAAGGTTCCCTTATGATGCATCATC 413
QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
Db 414 AGTGAAGAACTTCTCATGAAGATCCGAATTTATCAGTTCGAAGATTCAGACATTTGTCAT 473
QY 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysAsp 119
Db 474 GAAATCAAGGGGGGCTGATGTGATTAAAGGTAGACGGAATACTATGTTTACCTTAAAGAT 533
QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
Db 534 GCGGCCCATCGGACACATATTCGGACAAAGAAAGAGATTAAACGCTCAGAAAGCAGGAACAC 593

494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluThrGluAsp 513
 1623 CCGATTCGTTCATCCAGAACGTTTAGGAAAACCAATCGCAATTTACTACCTGATGAT 1682
 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrSerAspGlyTyrIlePheAsp 533
 1683 GAGATTCAAGTAGCAAGTTGGCAGGAGTACACACAGAGAGCGGTATATCTTTGAT 1742
 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553
 1743 CCTCGTATATACACAGTATGAGGGGATCGCTTATGTAACTCCACATATGACCATAGC 1802
 554 HisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr 573
 1803 CACTGGATTAAAAAAGATAGTTTCTCTCAAGCTGAGAGCGGCGAGCCGCTTATGCT 1862
 574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593
 1863 AAGAGAAAGGTTTGAACCCCTCTTCGACACACACAGGATTGAGAAATACTGAGGCA 1922
 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArg 613
 1923 AAGAGAGCAGAGCTATCTACAAACCGCTGAAAGCAGCTAAGAGAGTGCCTTATGATCGT 1982
 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633
 1983 ATGCTTCAATCTCAATATATCTACTGAGAGTCAAAACCGGTAGTTTATCTACTCAT 2042
 634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653
 2043 TATGACCAATACCATACATCAATTTGAGTGTTCGAGAGGCTTTATGAGGACCT 2102
 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673
 2103 AAGGGTATAGTCTTTCAGGATCTTTTGGCAGCTGTCAAGTACTATGTGCAACATCAAC 2162
 674 GluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLys 692
 2163 GAACCTCCGATTCAGATAATGTTTGTGTAACGCTAGTGCACCTGTTGTAATAAATAG 2222
 693 -----LysAspHisSerGlu----- 697
 2223 GCAGACCAAGATAGTAACTGATGAGATAAGGAACATGATGAAGTAAAGTGAAGCAACT 2282
 697 ----- 697
 2283 CACCCTGAATCTGATGAAAAAGAGAAATCAACGCTGTTTAAATCTTCACAGATAATCTT 2342
 698 ---AspProAsnLysAsnPhelysAlaAspGluGluProValGluGluThrProAlaGlu 716
 2343 TATAACCAAGCACTGATACGGAAGAGACAGAGGAGAGAGCTGAGATACCAAGATGAG 2402
 717 ProGluValProGlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluVal 736
 2403 GCTGAAATTCCTCAAGTAGAGAAATCTGTTTAACTGCTAAGTAGACAGATGCGAGGCC 2462
 737 LeuLeuAlaLysValThrAspSerLeuLysAlaAsnAlaThrGluThrLeuAlaGly 756
 2463 TTGCTAGAAAAAGTAAACAGATCTCTAGTATTAGACAAAATGCTATGAGACATTCGCTGT 2522
 757 LeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGlu 776
 2523 CTAAAAAGTAGTCTTCTCTCGGAACGAAAGATATTAACACTATTTTCAGCAGAGTAGAT 2582
 777 LysLeuLeuAlaLeuLysGlySerAsnProSerVal 790
 2583 AGTCTCTGCTTGTGTTTAAAGAAAGTCAACCGCTCCTATA 2624

RESULT 16

US-09-765-272-65
 ; Sequence 65, Application US/09765272
 ; Patent No. US20020061545A1
 ; GENERAL INFORMATION:

140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
 594 AGTCATAATCATAACTCA-----AGACGACATAATGCTGTGTCGACGACAGGCC 644
 160 GlnGlyArgTyrThrThrAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
 645 CAAGGAGCTTATACCAACGATGATGGGTATATCTTCATGCTATCTGATATCAATTGAGGAC 704
 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
 705 ACGGCTGATCTTATATGTTCTTCACGGCAGCATTACCATTAATCTTCAAGATGAG 764
 200 LeuSerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
 765 TTATCAGCTAGCAGTATAGCTGCTGCAGAACCTTATGGAATGG----- 809
 220 AsnSerArgThrTyrArgGluAsnSerAspAsnThrSerArgThrAsnTrpValPro 239
 810 -----AAGCAGGATCTCGTCTCTTCAAGTCTTAGTTATATGCA 851
 240 SerValSerAsnProGlyThrThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
 852 AATCCAGTTCAACCAAGATGTGCAGAAACCAACATCTCACTGTCTCACTCCCACTTATCAT 911
 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
 912 CAA-----AATCAAGGGGAAACATTTCAAGCCCTTTACGTGAAATGTATGCTTAAACCTTA 968
 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
 969 TCAGAACGCCATGTAGATCTGATGCGCTTATTTTCGACCCAGCGCAATCAACAAGTCA 1028
 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
 1029 ACCGCCAGAGGTAGTCTGCTCATGTGTAAACCATTAACCTTTATCCCTTATGCAACAA 1088
 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
 1089 ATGCTGGAATGAAAGCAAGATGTCTGATATTTCCCTTCCTGTTATGTTCAACCAT 1148
 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSer 358
 1149 TGGTACCAAGATCAAGACCAAGAACCAACCAAGTCCCAATCGACTCCGGAACCTAGTCCA 1208
 359 -----ProGlyProGlnProAlaProAsnLeuLysIleAspSerAsnSer 373
 1209 AGTCTGCAACCTGCACCAATCTCAACAGCTCCCAAGCAATCCAATTGAT-----GAG 1262
 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393
 1263 AAATGGTCAAAAGACTGTTTCAAAAGTAGGCGATGGTATGTTATCTTTGAGGAGATGGA 1322
 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413
 1323 GTTCTCTGTTATATCCAGCCAGGATCTTTCAGCAGAAACACAGCAGGAGCATTTGATAGC 1382
 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433
 1383 AAACCTGCCAAGCAGGAAAGTTTATCTCAAGCTAGGAGCTAAGAAACCTGACCTCCCA 1442
 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453
 1443 TCTAGTATCGAAGATTTTACATAGGCTTATGATCTTACTAGCAAGATTCACCAAGAT 1502
 454 LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473
 1503 TTACTTCAATAAAGTTCACCAAGTGTATTTGAGGTTTTTGGATATACCTGTTGGAACGA 1562
 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla 493
 1563 CTCAAGGATGTCTCAAGTGAATAAGTCAAGTTAGTGTATGATATTTCTGCTTCTTAGCT 1622

; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ;
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/765,272
 ; FILING DATE: 22-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ;
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ;
 ; INFORMATION FOR SEQ ID NO: 65:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2290 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ;
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 65:
 ; US-09-765-272-65

Alignment Scores:
 Pred. No.: 9.94e-231 Length: 2290
 Score: 2649.50 Matches: 519
 Percent Similarity: 76.51% Conservative: 77
 Best Local Similarity: 66.62% Mismatches: 130
 Query Match: 63.61% Indels: 53
 DB: 9 Gaps: 8
 ;
 US-09-765-271-56 (1-796) x US-09-765-272-65 (1-2290)
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19
 DB 5 TCCTATGAACCTTGGTCTGCTCACCAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTTCT 64
 QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
 DB 65 TATATAGATGGTGATCAGGCTGTGTCAAAAGCGCAGAAACTTGACACGATGAAGTCAGT 124
 QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
 DB 125 AAGAGGGAGGGGATCAACCGCGACCAATNGTNATCAAGATTACGGATCAAGTTATGTG 184
 QY 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIle 79
 DB 185 ACCTCTCATGGAGACCATATATCTACTATAATATGGCAAGGTTCTTATGATGCGCATCAIC 244
 QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
 DB 245 AGTGAAGAGCTCTCATGAAGAATCCGAATATACAGTTGAAGGATTCAGACATTCATCAAT 304
 QY 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
 DB 305 GAANTCAAGGGTGGTATGTCTATAGGTAACGGTAATATCTATGNTACCTTAAAGGAT 364
 QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139

DB 365 GCAGCTCATGGGATAATATTTCGGACAAAAGAGAGATTAAACGTCAGAAAGCAGAAACGC 424
 QY 140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
 DB 425 AGTCATATATCACTCA-----AGACAGATAATGCTGTGTCGACCCAGAGACC 475
 QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
 DB 476 CAAGGACGTTATACAAACGATGATGGGTATATCTTCAATGCATCTGATATCATTCAGGAC 535
 QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
 DB 536 ACGGTGATGCTTATATCGTTCTCCACGCGACCATTAACCATTAATCTCTAAGAAATGAG 595
 QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
 DB 596 TTATCAGCTACGAGTTAGCTGCTGCAGAACCTATTGGAATGGG----- 640
 QY 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239
 DB 641 -----AAGCAGGATCTCGTCTCTTCAAGTTCTAGTTATAATGCA 682
 QY 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
 DB 683 AATCCAGCTCAACCAAGATTGTCAGAAACCAATCTGACTGTCACCTCCAACTTATCAT 742
 QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
 DB 743 CAA---AATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTTGATGCTAAACCTTTA 799
 QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
 DB 800 TCAGACGCCATGTGGAATCTGATGGCTTATTTTCGACCAGCGCAATTCACAAGTCGA 859
 QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
 DB 860 ACCGCCAGAGGTGAGCTGCTCCCTCATGGTAACCATTACCACCTTATCCCTTATGAACAA 919
 QY 320 MetSerGluLeuGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
 DB 920 ATGCTGTAATGGAAAACCAATTTGCTGCTATTTATCCCTTCTGTTATCGTTTCAACCAT 979
 QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
 DB 980 TGGGTACCAATTCAAGACCCAGAACCAACCAAGTCCCAATCGACTCCGGAACCTAGTCCA 1039
 QY 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373
 DB 1040 AGTCCGCACTCCACCAATCTCAACAGCTCCCAAGCAATCCAATTGAT-----GAG 1093
 QY 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393
 DB 1094 AAATTGGTCAAGACCTGTTTCGAAAAGTAGGCGATGTTATGTTCTTTGAGGAGAATGGA 1153
 QY 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413
 DB 1154 GTTCTCTGTTATATCCCAAGGATCTTTCAGAGAAAACAGCAGCAGCATTTGATAGC 1213
 QY 414 LysLeuSerLysGlnGlnSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433
 DB 1214 AAATCGCCAGCAGGAAAGTTTATCTCATAGTAGGAGCTAAGAAACCTGACCTCCCA 1273
 QY 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453
 DB 1274 TCTAGTATCGAGATTTTACATTAAGGCTTATGACTTACTAGCAAGATTCCACCAAGAT 1333
 QY 454 LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473
 DB 1334 TTACTTGATAATAAAGGTCGACAAAGTTGATTTTGGAGCTTTGGATAACCTGTTGGACGA 1393
 QY 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspLeuLeuAlaPheLeuAla 493


```

COMPUTER READABLE FORM:
MEDIUM TYPE: CD-R
COMPUTER: Dell Latitude Pentium 3
OPERATING SYSTEM: Windows 98
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/158,844
FILING DATE: 03-Jun-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/961,527
FILING DATE: 1997-10-30
APPLICATION NUMBER: US 60/029,960
FILING DATE: 1996-10-31
ATTORNEY/AGENT INFORMATION:
NAME: Hyman, Mark J.
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB340P1D1
INFORMATION FOR SEQ ID NO: 243:
SEQUENCE CHARACTERISTICS:
LENGTH: 2359 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 243:
US-10-158-844-243

Alignment Scores:
Pred. No.: 4,02e-151 Length: 2359
Score: 1774.00 Matches: 347
Percent Similarity: 81.22% Conservative: 51
Best Local Similarity: 70.82% Mismatches: 74
Query Match: 42.59% Indels: 18
DB: 13 Gaps: 5

US-09-765-271-56 (1-796) x US-10-158-844-243 (1-2359)
QY 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnAsnArgValSer 19
Db 940 TCCTATGAGCTTGGACGTTTACCAAGCTGGTCAGGATAGAAAGAGTCTTAATCGAGTTGCT 999

QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
Db 1000 TATATAGATGGTGATCAGCTGGTCAAAGGCGAGAAAACCTTGACACCAGATGAAGTCACT 1059

QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
Db 1060 AAGAGGAGGGGATCAACGCCCAACAAATTTGTATCAAGATTACCGATACAGGTTATGTG 1119

QY 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79
Db 1120 ACCTCTCATGGAGACCATTTATCATTTATATATGCAAGGTTCCCTTATGATGCCATCATC 1179

QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
Db 1180 AGTGAGAGCTCTCATGAAGATCGGAATTTACGTTGAAGGATTCAGACATTTGTCAAT 1239

QY 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
Db 1240 GAAATCAAGGGTGGTTATGTCTCAATAGGTTAAACGGTAAATACTATGTTTACCTTAAAGT 1299

QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
Db 1300 GCAGCTCATCGGATTAATATTCGACAAAGAGAGATTTAAAGCTCAGAAGCAGGAACGC 1359

QY 140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
Db 1360 AGTCATAATCATAACTCA-----AGAGCAGATAATGCTGCTGTCGAGCCAGAGCC 1410

QY 160 GlnGlyArgTyrThrThrAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
Db 1411 CAAGGACGTTTATCAACGGATGGGTATATCTTCAATGCAATCTGATCATTTAGGAC 1470

```



```

; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; FILE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884, 465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 3120
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-1

Alignment Scores:
Pred. No.: 4,78e-103 Length: 3120
Score: 1247.00 Matches: 298
Percent Similarity: 54.36% Conserved: 95
Best Local Similarity: 41.22% Mismatches: 180
Query Match: 29.94% Indels: 150
DB: 10 Gaps: 19

US-09-765-271-56 (1-796) x US-09-884-465A-1 (1-3120)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db 64 GCCTATGCATTAACACAGCATGTTCCGAGGAAATAAGGACATATATCGTGTCTCTTAT 127
QY 21 ILAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 124 GTGGATGGCAGCCAGCTCAAGTCAGAAAAAGTGAACCTTGACACCAGCACCGTTTAGCCAG 183
QY 41 ArgGluGlyLleAsnAlaGluGlnIleValLleLysLleThrAspGlnGlyTyrValThr 60
Db 184 AAGAAGAGAAATTCAGCTCAGCAAAATGTAACTCAAAATATACAGATCAGGGCTATGTAAACG 243
QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaLleIleSer 80
Db 244 TCACACGGTGACCACATCATCTACTATATATCGGAAAGTTCCTTAATGATGCCCTCTTTAGT 303
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspLleValAsnGlu 100
Db 304 GAAGAACTCTTGATGAAGGATCCAAACTATCAACTTAAAGAGCGTGATATTGTCAATGAA 363
QY 101 ValLysGlyGlyTyrValLleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db 364 GTCAGAGGGGGTATATCATCAAGTCGATGGAATAATTATGTCTACCTGAAAGATGCA 423
QY 121 AlaHisAlaAspAsnValArgThrLysGluLleAsnArgGlnLysGlnGluHisSer 140
Db 424 GCTCATGCTGATAAATGTTCCGAACCTAAAGATGAATCAATCGTCAAAAAACAAGAACATGTC 483
QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
Db 484 AAGAATAATGAG- - - - -AAGGTTAACTCTAATGTTGCTGTAGCAAGGCTCTCAG 531
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspLleLleGluAspThr 180
Db 532 GGACGATATACAAATATGATGGGTATGCTTTTAATCCAGCTCATATTTATCGAAGATAG 591
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db 592 GGTAAATGCTTATATCGTTCCTCATGGAGGTCACTATCATCTACATTCCTCCAAAAGCGAATTA 651
QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db 652 TCTGTAGTAGTAATTAGCAGCAGCTAAAGCACATCTGGCTGGAAAAAATATGCAACCGAGT 711
QY 221 SerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
Db 712 CAGTTAAGCTATTCTTCAACAGCTAGTGACAAAT- - - - - 744

```

[illegible]

Qy 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnThrAsnSerGln 260
 Db 745 -----AACACGCAATCTCTAGCAAAAGGATCAACATAGCAAGCCA 783
 Qy 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
 Db 784 GCAATAAATCTGAATATCTCAGAGCTTTTGAAGGAACTATGATTCACCTAGCGCC 843
 Qy 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
 Db 844 CAACGTTACAGTGAATCAGATGGCTGCTTTTACCCCTGCTGAAGATATCAGTCGTACA 903
 Qy 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
 Db 904 CCAATGAGTGGATGCGATCCCGATGGCCACCATACCATTTATTCCTTACAGCAAGCTT 963
 Qy 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340
 Db 964 TCTGCTTTAGAGAAAGATTTGCCAGATGGTGGCT----- 999
 Qy 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
 Db 999 ----- 999
 Qy 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
 Db 1000 -----ATCAGTGGAACTGGTCTTACAGTCTTCTACAAATGCA 1035
 Qy 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
 Db 1036 AAA----- 1038
 Qy 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
 Db 1039 -----CCTAATGAAGTACTGTCTAGTCTAGGCGAGTCTTTCAAGCAATCTTCTCT 1089
 Qy 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
 Db 1090 -----TTACGACAAAGTAAAGAGCTCTCTTCAGATCTGCTGTTATTTT 1137
 Qy 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
 Db 1138 AATCCAAA---GATATCGTTGAAGAAACGGCTACAGCTTATATTGTAAGACATGCT--- 1191
 Qy 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
 Db 1192 ---GATCAATTCCATTACATTCACAAA-----TCAAAATCAAAATGGG 1230
 Qy 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
 Db 1231 CAACCGACTCTCCAAACAAATAGCTAGCAACACCTTCTCCA---TCTCTCCAAATCAAT 1287
 Qy 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
 Db 1288 CCAGGAACCTTCATCAGAAACATGAA----- 1314
 Qy 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
 Db 1315 -----GAAAGATGGATACGGATTTGATGCTAATCTATATTCGCTGAA 1356
 Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560
 Db 1357 GATGAATCAGGTTTGTTCATGAGTACGGAGACCAATCATTTATTTCTTCAGAGAGGAC 1416
 Qy 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
 Db 1417 TTGACAGAGAGCAAAATTAAGCTGGCGCAAAAACATTTAGAGGAA----- 1461
 Qy 581 ProSerProAspAlaAspValLysAlaAsnProThrGly---AspSerAlaAlaAlaIle 599
 Db 1462 -----GTTAAACTAGTACATAATGATTTAGTATCTTCTGTCATCTCAT 1503

Qy 600 TyrAsnArgValLysGly-----GluLys 607
 Db 1504 GAACAGGATTATCCAGGTAATGCCAAAGAAATGAAGATTTAGATAAAAAAATCGAAGAA 1563
 Qy 608 ArgIleProLeuValArgLeuProTyrMetVal-----GluHisThrValGluValLys 625
 Db 1564 AAAATCTGCTGCAATATGAACAATATGGTGTCAACAGTGAAGATTTGTCGTAATAAA 1623
 Qy 626 AsnGlyAsnLeuIleIle---ProHisLysAspHisTyrHisAsnIleLysPheAlaTrp 644
 Db 1624 GAAAAAATCGATATTTATCCGATGGAGATCACCATCATGCAGATCCG----- 1674
 Qy 645 PheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAlaThr 664
 Db 1675 ATTGATGAACAT-----AAACCGTGTGGAATGGT-----CATTCCTAC 1713
 Qy 665 IleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAspGly----- 681
 Db 1714 AGTAACATAGTAACATGTTTAAACCCGAGAGAGAGTGTCTAAAAAGAGAGGAATAAGTT 1773
 Qy 682 ---TrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
 Db 1774 TATACTGGAGAGAAATTAACGAATGTTGTTAATTTGTTAAAAAATAGTACGTTTAATA 1833
 Qy 701 LysAsnPhe 703
 Db 1834 CAAAACTTT 1842

RESULT 19
 US-09-884-465A-2
 ; Sequence 2, Application US/09884465A
 ; Publication No. US20030077293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis
 ; APPLICANT: Charland, Nathalie
 ; APPLICANT: Ouellet, Catherine
 ; TITLE OF INVENTION: Streptococcus Antigens
 ; FILE REFERENCE: 055190-0044
 ; CURRENT APPLICATION NUMBER: US/09/884,465A
 ; CURRENT FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: 60/212,683
 ; PRIOR FILING DATE: 2000-06-20
 ; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 2
 ; LENGTH: 5048
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-09-884-465A-2

Alignment Scores:
 Pred. No.: 9,27e-103 Length: 5048
 Score: 1247.00 Matches: 298
 Percent Similarity: 54.36% Conservative: 95
 Best Local Similarity: 41.22% Mismatches: 180
 Query Match: 29.94% Indels: 150
 DB: 19 Gaps: 19

US-09-765-271-56 (1-796) x US-09-884-465A-2 (1-5048)

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
 Db 1840 GCCTATGCACTAAACAGCATCGTTCGAGGAAAAATAAGCAATAATCGTGTCTCTAT 1899
 Qy 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
 Db 1900 GTGGATGGCAGCAGTCAAGTCAGAAAGTGAAGAACTTGACACAGCAGGTAGCCAG 1959
 Qy 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
 Db -----

Db 1960 AAAGAGGAATTCAGCGTACGAAATTTGTAATCAAAATTTACAGATCAGGGCTATGTAACG 2019

QY 61 SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer 80

Db 2020 TCACAGGTGACCATATCATTTATATGTTGGAAGTTCTTATGATGCCCTCTTTAGT 2079

QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100

Db 2080 GAAGAAGCTCTTGATGAAGATCCAACTATCACTTAAAGAGCGTGATATTGTCATGAA 2139

QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysAspAla 120

Db 2140 GTCAGGGTGGTTATCATCAAGTCGATGAAATATTTATGTCAGCTGAAGATGCA 2199

QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGluHisSer 140

Db 2200 GCTCATGCTGATATGTTGCAATAGATGAATCAATCGTCAAAACAAAGACATGTC 2259

QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160

Db 2260 AAAGATAATGAG-----AAGGTTAACTCTAATGTTGCTGAGCAAGGTCTCAG 2307

QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180

Db 2308 GGACGATATAGCAAAATGATGTTATGTTCTTAATCCAGCTGATATTATCGAAGATACG 2367

QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200

Db 2368 GGTAATGCTTATATCGTTCTTCATGAGGTCACCTATCACTACATCTCCAAAGCGATTGA 2427

QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220

Db 2428 TCTGCTAGTGAATTAGCAGCAGCTAAAGCACATCTGGCTGGAAAAATATGCAACCGAGT 2487

QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240

Db 2488 CAGTTAGCTATTCTTCAACAGCTAGTGACAAAT----- 2520

QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260

Db 2521 -----ACACGCAATCTGTAGCAAAAGGATCAACTAGCAAGCCCA 2559

QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLysGlnLeuTyrLysLeuProLeuSer 280

Db 2560 GCAAAATAAATCTGAAATCTCCAGAGCTTTTGAAGGAACCTCTATGATTCACCTAGGCC 2619

QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300

Db 2620 CAACGTTACAGTGAATCAGATGGCCCTGGTCTTTGACCCCTGCTAAGATATTATCAGTCGTACA 2679

QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320

Db 2680 CCAAAATGGAGTTGCGATTCGCGATGCGACCATTAACCTTTATTCCTTACGAAGCTT 2739

QY 321 SerGluLeuGluGluArgIleAlaArgIleLeuProLeuArgTyrArgSerAsnHisTrp 340

Db 2740 TCTGCTTTAGAGAAAGATGTCAGATGGTGCCT----- 2775

QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360

Db 2775 ----- 2775

QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380

Db 2776 -----ATCAGTGAACGTGTTCTACAGTTTCTCAAAATGCA 2811

QY 381 ArgLysValGlyGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400

Db 2812 AAA----- 2814

QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420

Db 2815 -----CCTAATGAAGTAGTGTCTAGCTAGGCAAGTCTTTCAAGCAATCCTTCTTCT 2865

QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440

Db 2866 -----TTAAACGACAGTAAGGAGCTCTCTTCAGCATCTGATGGTTATATTTT 2913

QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460

Db 2914 AATCCAAA-----GATATCGTTGAAGAAACGGCTACAGCTTATATTGTAAGACATGCT-- 2967

QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480

Db 2968 ---GATCATTTCCATTACATTCCTCAAA-----TCAAAATCAAAATCGG 3006

QY 481 LysGluLysLeuValAspAspLeuAlaPheLeuAlaProIleThrHisProGluArg 500

Db 3007 CAACCGACTCTTCCAAACAATAGTCTAGCAACACCTTCTCCA---TCTCTTCCAATCAAT 3063

QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520

Db 3064 CCAGGAACCTTCACATGAGAAACATGAA----- 3090

QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540

Db 3091 -----GAAGATGGATACGATTTGATGCTAATCTATTTATTCGCTGAA 3132

QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560

Db 3133 GATGAATCAGGTTTGTGTCATGAGTCACGGAGACCAACATCATTTATTTCTCAAGAGAC 3192

QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580

Db 3193 TTGACAGAGAGCAAAATTAAGGCTGCGCAAAACATTTAGAGGAA----- 3237

QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGly---AspSerAlaAlaAlaIle 599

Db 3238 -----GTTAAACATGATGATAATGATGATTCTTTGTCTATCTCAT 3279

QY 600 TyrAsnArgValLysGly-----GluLys 607

Db 3280 GAACAGGATTTATCCAGTAAATGCAAGAAATGAAGATTTAGATAAAAAATCGAAGAA 3339

QY 608 ArgIleProLeuValArgLeuProTyrMetVal-----GluHisThrValGluValLys 625

Db 3340 AAAATTTGCTGGCAATTAATAACAATATGTTGTCAAAACGTCAAAAGTATTGTCGTGAATAAA 3399

QY 626 AsnGlyAsnLeuIleIle---ProHisLysAspHisTyrHisAsnIleLysPheAlaTrp 644

Db 3400 GAAAAAATCGCATTTATTTATCCGCATGGAGATCACCATCATCGACATCCG----- 3450

QY 645 PheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAlaThr 664

Db 3451 ATTGATGAACAT-----AAACCGTTGGAATTTGGT-----CATTTCTCAC 3489

QY 665 IleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAspGly----- 681

Db 3490 AGTAATGTAACACTGTTTAAACCCGGAAGAGGATTTGCTTAAAAAAGAGGAATAAAGTT 3549

QY 682 ---TrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700

Db 3550 TATATCTGGAGACAATTAACGATTTGTTAATTTGTTAAAAAATAGTAGCTTTAATAAT 3609

QY 701 LysAsnPhe 703

Db 3610 CAAAACCTTT 3618

RESULT 20

US-10-158-844-192/c

; Sequence 192, Application US/10158844

; Publication No. US20040029118A1

; GENERAL INFORMATION:

; APPLICANT: Kunsch et al.

; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 391

Fri Oct 1 15:35:52 2004

CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: CD-R
COMPUTER: Dell Latitude Pentium 3
OPERATING SYSTEM: Windows 98
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/158,844
FILING DATE: 03-Jun-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/961,527
FILING DATE: 1997-10-30
APPLICATION NUMBER: US 60/029,960
FILING DATE: 1996-10-31

ATTORNEY/AGENT INFORMATION:
NAME: Hyman, Mark J.
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB340PID1

INFORMATION FOR SEQ ID NO: 192:
SEQUENCE CHARACTERISTICS:
LENGTH: 6867 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 192:
US-10-158-844-192

Alignment Scores:
Pred. No.: 2,39e-100 Length: 6867
Score: 1222.50 Matches: 295
Percent Similarity: 54.22% Conservative: 97
Best Local Similarity: 40.80% Mismatches: 181
Query Match: 29.35% Indels: 151
DB: 13 Gaps: 18

US-09-765-271-56 (1-796) x US-10-158-844-192 (1-6867)

QY	1	SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr	20
Db	6722	GCCTATGCCTAAACACGATCGTTCGAGAAATAAGCAACAATATCGTCTCTTAT	6663
QY	21	IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys	40
Db	6662	GTGGATGGCAGCCAGTCAAGTCAAGAAAGTGAAACTTGACACACAGCAGGTAGCCAG	6603
QY	41	ArgGluGlyIleAsnAlaGluGlnIleValLysIleThrAspGlnGlyTyrValThr	60
Db	6602	AAAGAGGATTCAGCTGAGCAATATGTAATCAAAATACAGATCAGGCTATGTAACG	6543
QY	61	SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer	80
Db	6542	TCACAGGTCACCATCATCTATTAATGGAAGAGTTCCTATGATGCCCTCTTTAGT	6483
QY	81	GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu	100
Db	6482	GAGAAGACTTGTGATGAGGATCCAACTATCAACTTAAGACGCTGATATGTCATGAA	6423
QY	101	ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla	120
Db	6422	GTCAGGGTGGTATATCATCAAGTGCATGGAATAATATTATGTCATCTGAAAGATGCA	6363
QY	121	AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnHisSer	140
Db	6362	GCTCATGCTGATAATGTTCCAACTAAAGATCAATCGTCAAAACAGAACATGTC	6303
QY	141	GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln	160

Db	6302	AAAGATAATGAG-----AAGTTAACTCTAAGTTCTCTAGCAAGTCTCAG	6255
QY	161	GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr	180
Db	6254	GGACGATATACGACAAATGATGTTATCTTTAAATCCAGCTGATATTATCGAAGATACG	6195
QY	181	GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu	200
Db	6194	GGTAATGCTTATATCGTTCTCATGAGGTCACATCATCATCATCTCCAAAGAGATTTA	6135
QY	201	SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn	220
Db	6134	TCTGCTAGTGAATTAGCAGCAGCTAAAGCACATCTGCTGGAAAAATAATGCCACCGAGT	6075
QY	221	SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer	240
Db	6074	CAGTTAAGCTATTCTTCAACAGCTAGTGACAAAT-----	6042
QY	241	ValSerAsnProGlyThrThrAsnThrSerAsnAsnSerAsnThrAsnSerGln	260
Db	6041	-----AACACGCAATCTCTAGCAAAAGATCAACTAGCAAGCCCA	6003
QY	261	AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer	280
Db	6002	GCAAAATAATCTGAAATCTCCAGAGTCTTTTGAGGAACTCTATGATTCACCTAGCGCC	5943
QY	281	GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr	300
Db	5942	CACGTTACAGTGAATCAGATGGCTGCTCTTACCCCTGCTAAGATTATCAGTCGTACA	5883
QY	301	AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet	320
Db	5882	CCAAATGAGATTGCGATTCGACATGCGACCATACCACTTTATCTTACAGCAAGCTT	5823
QY	321	SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp	340
Db	5822	TCTGCTTTAGAGAAAGATTGCCAGAAATGGTGCCT-----	5787
QY	341	ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly	360
Db	5787	-----	5787
QY	361	ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal	380
Db	5786	-----ATCAGTGGAACTGGTCTACAGTTCTCAAAATGCA	5751
QY	381	ArgLysValGlyGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla	400
Db	5750	AAA-----	5748
QY	401	LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer	420
Db	5747	-----CCATATGAAGTAGTGTCTAGTCTAGGCAGTCTTTCAAGCAATCTCTCTCT	5697
QY	421	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	440
Db	5696	-----TTAACGACAAAGTAAGAGCTCTCTCAGCATCTGATGTTATATTTT	5649
QY	441	AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg	460
Db	5648	AAATCCAAAA---GATATCGTTGAGAAACCGCTACAGCTATATTTGTAAGACATGGT---	5595
QY	461	AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn	480
Db	5594	---GATCAATTTCCATTTACATTTCCAAAA-----TCAAAATCAAAATGGG	5556
QY	481	LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg	500
Db	5555	CAACCGACTCTTCCAAACAAATAGTCTAGCAACACCTTCTCCA---TCTCTTCAATCAAT	5499
QY	501	LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu	520

Db 5498 CCAGGACTTCACATGAGAAACATGAA----- 5472

QY 521 AlaAspLysThrThrSerAspGlyTyrIlePheAspGluHisAspIleLeuSerAsp 540

Db 5471 -----GAAATGATGATGAGTTGATGCTAATCGTATTATCGCTGAA 5430

QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisThrIleGlyLysAspSer 560

Db 5429 GATGAATCAGGTTTGTTCATGAGTCAGCGAGACCAACATCATTTCTTCAAGAGGAC 5370

QY 561 LeuSerAspLysGluLysValAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580

Db 5369 TTGACAGAGAGCAAAATTA--GSTCGCAAAACATTTAGAGAA----- 5326

QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600

Db 5325 -----GTTAAACTAGTCAATGATGATTTGTCATCTCAT 5284

QY 601 AsnArg-----ValLysGlyLys 607

Db 5283 GAACAGGATTATCCAAAGTAATGCCAAAGAAATGAAGATTAGATAAAATAATCGAGAA 5224

QY 608 ArgIleProLeuValArgLeuProTyrMetVal-----GluHisThrValGluValLys 625

Db 5223 AAAATGCTGCATTATGAACAATATGCTGCAACGTAAGATGATTGCTGGAATAAA 5164

QY 626 AsnGlyAsnLeuIleIle---ProHisLysAspHisTyrHisAsnIleLysPheAlaTrp 644

Db 5163 GAAAAAATGCGATTATTTATCCGATGAGATCACCATCATGCGATCCG----- 5113

QY 645 PheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAlaThr 664

Db 5112 ATTGATCAACAT-----AAACCGGTTGCAATTGCT-----CATTCAC 5074

QY 665 IleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAspGly----- 681

Db 5073 AGTAACTATGACTGTTTAAACCGAAGAGAGCTTCTGCTAAAAAGAGGAATAAAGTT 5014

QY 682 ---TrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700

Db 5013 TATACTGGAGAGAATTAACGAATGTTGTTAATTTGTTAAATAATAGTACGTTTAATAAT 4954

QY 701 LysAsnPhe 703

Db 4953 CAAAACTTT 4945

RESULT 21

US-09-769-787-246

; Sequence 246, Application US/09769787

; Publication No. US20030091577A1

; GENERAL INFORMATION:

; APPLICANT: Microbial Technics Limited

; APPLICANT: Gilbert, Christophe FG

; APPLICANT: Hansbro, Philip M

; TITLE OF INVENTION: Proteins

; FILE REFERENCE: FWC/P21129WO

; CURRENT APPLICATION NUMBER: US/09/769,787

; PRIOR FILING DATE: 2001-01-26

; PRIOR APPLICATION NUMBER: GB 9816337.1

; PRIOR FILING DATE: 1998-03-27

; PRIOR APPLICATION NUMBER: US 60/125164

; PRIOR FILING DATE: 1999-03-19

; NUMBER OF SEQ ID NOS: 388

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 246

; LENGTH: 1455

; TYPE: DNA

; ORGANISM: Streptococcus pneumoniae

US-09-769-787-246

Alignment Scores:

Pred. No.: 1,51e-99 Length: 1455

Score: 1203.50 Matches: 259

Percent Similarity: 59.47% Conservative: 77

Best Local Similarity: 45.84% Mismatches: 120

Query Match: 28.90% Indels: 109

DB: 10 Gaps: 10

US-09-765-271-56 (1-796) x US-09-769-787-246 (1-1455)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnArgValSerTyr 20

Db 64 GCCTATGCACTAAACCCAGCATCGTTTCGAGGAAATAGGACAAATATCGTCTCTTAT 123

QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40

Db 124 GTGGATGCGACCCAGTCAAGTCAGAAAAGTGAACACTTGACACAGACCGGTTAGCCAG 183

QY 41 ArgGluGlyIleAsnAlaGluLysIleValLysIleThrAspGlnGlyTyrValThr 60

Db 184 AAGAGGAATTCAGGCTGAGCAAAATTTGATCAAAATTACAGATCAGGGCTATGTAACG 243

QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleSer 80

Db 244 TCACACGGTGACCACTATCATTAATAATGSGAAAGTTCTCTATGATGCCCTCTTTAGT 303

QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLysLysAspGluAspIleValAsnGlu 100

Db 304 GAAGAACTCTTGATGAAGGATCCAAACTATCAACTTAAAGACGCTGATATTGTCAATGAA 363

QY 101 ValLysGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120

Db 364 GTCAAGGGTGGTTATATCATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 423

QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140

Db 424 GCTCATGCTGATAATGTTGCAACTAAAGATGAAATCAATCGTCAAAACACAGAACATGTC 483

QY 141 GlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160

Db 484 AAGATAATAG-----AAGTTAACTCTAATGTTGCTGAGCAAGTCTCAG 531

QY 161 GlyArgTyrThrThrAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180

Db 532 GGACGATATACGACAATGATGTTATGCTTTAATCCAGCTGATATTATCGAAGATACG 591

QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200

Db 592 GGTAATGCTTATATCGTCTCATGAGGTCACATCACTACATTCCTCAAAAGCGATTGA 651

QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220

Db 652 TCTGCTAGTGAATTAGCAGCAGCTAAAGCACATCTGCTGGAATAATATGCAACCGAGT 711

QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240

Db 712 CAGTTAAGCTATCTTCAACAGCTAGTGCAAT----- 744

QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260

Db 745 -----AACCGCAATCTGTAGCAAAAGGATCACTAGCAAGCCCA 783

QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLysGlnLeuTyrLysLeuProLeuSer 280

Db 784 GCAAAATAATCTGAAAATCTCCAGAGTCTTTTGAAGGAACCTCTATGATTCACCTAGCGCC 843

QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300

Db 844 CAACGTTACAGTAATCAGATGCGCTGCTTTGACCTGCTGCTGCTGCTGCTGCTGCTGCT 903

QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320

Db 904 CCAAAATGAGTGGATTCGCGATGCGGACCACTTACCACTTTATTCCTTACGCAAGCTT 963

QY 321 SerGluLeuGluArgIleAlaArgIleLeuProLeuArgTyrArgSerAsnHisTrp 340

```

; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; US-09-769-744A-23

Alignment Scores: 1.51e-99 Length: 1455
Pred. No.: 1203.50 Matches: 259
Score: 59.47% Conservative: 77
Best Similarity: 45.84% Mismatches: 120
Best Local Similarity: 28.90% Indels: 109
Query Match: 10 Gaps: 10
DB:

US-09-765-271-56 (1-796) x US-09-769-744A-23 (1-1455)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db 64 GCCTATGCATTAACACGAGCATCGTTCAGGAAAAATAAGGCAATAATCGTGTCTCTTAT 123
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 124 GTGGATGCGACCCAGTCAGTCAAGTCAGAAAGTGAAACCTTGACACGAGCCAGGTAGCCAG 183
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db 184 AAAGAGGAAATTCAGGCTGACCAATTCGTAATCAAATTCAGATCAGGGCTATGTAACG 243
QY 61 SerHisGlyAspHisTyrTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db 244 TCACACGGTGACCACTATCATTACTATAATGGAAAGTTCCTTATGATGCGCTCTTATG 303
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db 304 GAAGAACTCTTCATGAGGATCCAACTATCAACTTAAAGACGCTGATATTGTCATGAA 363
QY 101 ValLysGlyGlyTyrValLysValAspGlyLysTyrTyrValTyrValLysAspAla 120
Db 364 GTCAAGGGTGGTATTATCATCAAGTCGATGGAATATATTATGCTACCTGAAGATGCA 423
QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnIleHisSer 140
Db 424 GCTCATGCTGATATATGTCGAACTAAAGATGAATCAATCGTCAAAAAACAAGATGTC 483
QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaAlaArgSerGln 160
Db 484 AAAGATAATGAG-----AAGGTAACTCTAATGTTGCTGTACCAAGGCTCAG 531
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db 532 GGACGATATACGACAAATGATGTTATGCTTTTATCCAGCTGATATTATCGAAGATACG 591
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db 592 GGTAATGCTTATATCGTTCCTCATGGAGGTCACTATCACTACATTCCTCAAAAACGCAATTA 651
QY 201 SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db 652 TCTGCTAGTGAATTAGCAGCAGCTAAAGCACATCTCGCTGGGAAAAAATATGCAACCGAGT 711
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
Db 712 CAGTTAAGCTATTCTTCAACAGCTAGTGACAAAT-----744
QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db 745 -----AACACGCAATCTGTAGCAAAAGATCACTAGCAACCCA 783
QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLysGlnLeuLysLeuProLeuSer 280
Db 784 GCAAAATAATCTGAAATCTCCAGAGTCTTTTGAAGGAATCTCTATGATTACCTAGCGCC 843
QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db 844 CAACGTTACAGTGAATCAGATGGCTCTTTGACCCCTGCTAAGATATATCAGTCGTACA 903

```

```

964 TCTGCCTTAGAAGAAAGATTGCCAGATGCTGCT-----999
341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
999 -----999
361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
1000 -----ATCAGTGGAACTCGTTCACAGCTTCTTACAAATGCA 1035
381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
1036 AAA-----1038
401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
1039 -----CCTAATGAAGTAGTCTAGTCTAGGCACTCTTCAAGCAATCTCTTCT 1089
421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
1090 -----TTAACGACAAAGTAAGGAGCTCTTTCAGCATCTGATGGTTATTTT 1137
441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
1138 AATCCAAA---GATATCGTTGAGAAACCGCTACAGCTTATATTGTAAGACATGCT---1191
461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
1192 ---GATCATTTCATTACATTCCAAAA-----TCAAAATCAAAATTCGG 1230
481 LysGluLysLeuValAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
1231 CAACCGACTCTTCCAAACAATAGTCTAGCAACACCTCTTCCA---TCTCTTCCAATCAAT 1287
501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
1288 CCAGAACTTCATCAGTCAAGAAATGCA-----1314
521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
1315 -----GAAGATGATACCGATTGATGCTAATCGTATTATCGCTGAA 1356
541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560
1357 GATGAATCAGGTTTGTCTGATGATGCTACGAGACCAACATCAATATTCTTCAGAGAGAC 1416
561 LeuSerAspLysGlu 565
1417 TTGACAGAGAGCAA 1431

```

```

RESULT 22
US-09-769-744A-23
; Sequence 23, Application US/09769744A
; Publication No. US20030134407A1
; GENERAL INFORMATION:
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hanniffy, Sean B
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 1455

```

```

QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
904 CCAAAATGGAGTGGCATTCGCGATGCGGACCACTTACCACTTTATCTTACGCAAGCTT 963
QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
964 TCTGCTTTAGAAAGAAAGATTCCGAATGGTGCT----- 999
QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db ----- 999
QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1000 -----ATCAGTGGCACTGGTCTTACAGTTTCTACAAATGCA 1035
QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1036 AAA----- 1038
QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1039 -----CCTAATGAAGTAGTCTCTAGCTAGCGAGTCTTCAAGCAATCCTTCTCT 1089
QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1090 -----TTAACGACAAAGTAAGGAGTCTCTTCAGCATCTGATGTTATATTTT 1137
QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1138 AATCCAAAA---GATATCGTTGAAGAAACGCTACAGCTTATATTGTAAGACATGGT--- 1191
QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1192 ---GATCATTTCCATTACATTCGAAA-----TCAAAATCAAAATGGG 1230
QY 481 LysGluLysLeuValAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1231 CAACCGACTCTCCAAACAATAGTCTAGCAACACCTTCTCCA---TCTCTCCAAATCAAT 1287
QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1288 CCAAGAACTTCACATCAGAAACATGAA----- 1314
QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1315 -----GAAGATGATACGGATTGATGCTTAATCGTATTATTCGCTGAA 1356
QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1357 GATGAATCAGGTTTGTCTATGATGATCAGCGAGACCAACATATTATTCTTCAGAAGGAC 1416
QY 561 LeuSerAspLysGlu 565
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1417 TTGACAGAAGACAA 1431

```

RESULT 23

```

US-10-412-862-7
; Sequence 7, Application US/10412862
; Publication No. US20040052781A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; FILE OF INVENTION: Motifs
; FILE REFERENCE: 469201-685
; CURRENT APPLICATION NUMBER: US/10/412,862
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048

```

```

; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 7
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-412-862-7

```

Alignment Scores:

```

Pred. No.: 1,51e-99 Length: 1455
Score: 1203.50 Matches: 259
Percent Similarity: 59.47% Conservative: 77
Best Local Similarity: 45.84% Mismatches: 120
Query Match: 28.90% Indels: 109
DB: 13 Gaps: 10

```

```
US-09-765-271-56 (1-796) x US-10-412-862-7 (1-1455)
```

```

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
64 GCCTATGCACTAAACACGACATCGTTCCGAGAAATAAGACATAATCGTCTCTTAT 123
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
124 GTGATGGCAGCCAGTCAGTCAAGTCAGAAAGTGAACCTTGACACGACAGGTTAGCCAG 183
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
184 AAGAAGGAATTCAGGCTGAGCAAAATGTAATCAAAATACAGATCAGGCTATGTAACG 243
QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
244 TCACACGGTCACCACTATCTACTATATATGCGAAGTTCCTTATGTCCTCTTTAGT 303
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluValAsnGlu 100
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
304 GAAGAACTCTTGATGAAGGATCCAAACTATCAACTTAAAGACGCTGATATTGCAATGAA 363
QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
364 GTCAAGGGTGGTTATATCATCAAGGTCGATGGAATAATATTATCTACCTGAAAGATGCA 423
QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
424 GCTCATGCTGATAATGTTCCGAATAAGATGAATCAATCGTCAAAAACAAGAACATGTC 483
QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
484 AAGATAATGAG-----AAGGTTAACTCTAATGTTGCTAGCAAGGCTCTCAG 531
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
532 GCACGATATACGCAAAATGATGTTATGCTTTAATCCAGCTGATATTATTCGAAGATACG 591
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
592 GGTAATGCTTATATCGTCTCATGAGGTCACTATCACTATATCCCAAAAACCGATTTA 651
QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
652 TCTGCTAGTGAATTAGCAGCAGCTAAAGCAGCATCTGGCTGGAATAATATGCAACCGAGT 711
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
712 CAGTTAAGCTATTCTTCAACAGCTTAGTGACAA----- 744
QY 241 ValSerAsnProGlyThrThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
745 -----AACACGCAATCTGTAGCAAAAGGATCACTACCTAGCAAGCCA 783
QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

784 GCAATTAATCTGAAATCTCCAGAGCTTTTGAAGGAACCTCTATGATTCACCTAGCGCC 843
281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
844 CAAGTTACAGTGAATCAGATGCGCTGCTTTGACCCCTGCTAAGATATACAGTCGTACA 903
301 AlaArgGlyValAlaValProHisGlyAspHisThrPheIleProTyrSerGlnMet 320
904 CCAATGGAGTGGATGCGCATGCGCATGGGACCAITACCACCTTTATCTCTACAGCAAGCTT 963
321 SerGluLeuGluGluArgGlnAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340
964 TCTCCCTTAGAGAAAGATGCGCATGGGACCT----- 999
341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
999 ----- 999
361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
1000 -----ATCAGTGGAACTGGTCTCTACAGTTCTTACAAATGCA 1035
381 ArgLysValGlyGluGlyTyrValPheGluLysGlyIleSerArgTyrValPheAla 400
1036 AAA----- 1038
401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
1039 -----CCTAATGAAGTAGTGTCTAGCTAGGACGCTCTTCAAGCAAACTCTTCTCT 1089
421 ValSerHisThrLeuAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
1090 -----TTAACGCAAGATGAGGAGCTCTCTCAGCATCTGATGTTATATTTT 1137
441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
1138 AATCAAAA---GATATCGTTGAAGAAACGGCTACAGCTTATATTGTAAACATGCT--- 1191
461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
1192 ---GATCAATTCATTCATTCACAAATCCAAA-----TCAATCAATTTGGG 1230
481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
1231 CAACCGACTCTTCCAAACAATAGCTAGCAACACCTCTCCA---TCTCTTCAATCAAT 1287
501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
1288 CCAGGAACTTCACATGAGAAACATGAA----- 1314
521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
1315 -----GAAGATGGATACGGATTTGATGCTATATTCGTATATTCGCTGAA 1356
541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560
1357 GATGAATCAGGTTTGTGTCATGAGTCACGGAGACCAATCATTTATTCITCAAGAGGAC 1416
561 LeuSerAspLysGlu 565
1417 TTGACAGAGAGCA 1431

```

```

RESULT 24
US-10-412-850-7
; Sequence 7, Application US/10412850
; Publication No. US20040001836A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-696

```

```

; CURRENT APPLICATION NUMBER: US/10/412,850
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 3.0
; SEQ ID NO 7
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; US-10-412-850-7

Alignment Scores:
Pred No.: 1,51e-99 Length: 1455
Score: 1203.50 Matches: 259
Percent Similarity: 59.47% Conservative: 77
Best Local Similarity: 45.84% Mismatches: 120
Query Match: 28.90% Indels: 109
DB: 16 Gaps: 10

US-09-765-271-56 (1-796) x US-10-412-850-7 (1-1455)
QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
DB 64 GCCTATGACCTAAACCCAGCATCGTTCCGAGGAAATAAGGACAAATATATCGTCTCTTAT 123
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
DB 124 GTGATGGCAGCCAGTCAAGTCAGAAAGTGAAGAACTTGACACACAGCAGGTAGCCGA 183
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValLysIleThrAspGlnGlyTyrValThr 60
DB 184 AAAGAGGAATTCAGGCTGAGCAAAATTCATAAATTAACAGATCAGGGCTATGTAAAG 243
QY 61 SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
DB 244 TCACACGGTGACCATCTATCTATCTATATGGAAGATTTCTTATGATGCCCTCTTAGT 303
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysLysAspGluAspIleValAsnGlu 100
DB 304 GAAGAACTCTTGATGAGGATCCAACTATCAACTTAAAGACGCTGATATGTCAATGAA 363
QY 101 ValLysGlyGlyTyrValLysValAspGlyLysTyrTyrValTyrValTyrLysAspAla 120
DB 364 GTCAAGGGTGGTTATATCATCATCAAGTCGATGGAATAATATTATCTTACCTGAAAGATCA 423
QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
DB 424 GCTCATGCTGATAATGTTCCGAACATAAGATGAATATCNATCGTCNAAAACAAGACATGTC 483
QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
DB 484 AAAGATAATGAG-----AAGGTTAACTCTAATGTTGCTGTAGCAAGGCTCTCAG 531
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
DB 532 GGACGATATACGACAAATGATGTTATGTTTAAATCCAGCTGATATATTCGAAGATACG 591
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
DB 592 GGTAATGCTTATATCTCTCTCTCAAGGATGATCACTACATCTCCAAAAGCGGATTA 651
QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
DB 652 TCTGCTAGTAGAATTAGCAGCAGCTAAAGCAGCATCTGGCTGGAATAAATATGCAACCGAGT 711
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
DB 712 CAGTTAAGCTATTTCTCAACAGCGTAGTGACAAAT----- 744
QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260

```


Fri Oct 1 15:35:52 2004

Search completed: October 1, 2004, 13:50:16
Job time : 1107 secs

```

QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
Db :|||
712 CAGTTAAGCTATTCTTCAACAGCTAGTACAAAT----- 744
QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db :|||
745 -----AACCCCAATCTGTAGCAAAAGGATCAACTAGCAAGCCA 783
QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
Db :|||
784 GCAATAAATCTGAATCTCCAGAGCTTTTGAAGGAACCTCTATGATTCACCTAGCGGCC 843
QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db :|||
844 CAACGTTACAGTGAATCAGATGGCTGCTTTCCACCTGTCAAGATTATCAGTCGTACA 903
QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db :|||
904 CCAATGGAGTTGGATTCCGATGGCGACCATTACCACCTTATTCCTTACAGCAAGCTT 963
QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340
Db :|||
964 TCTGCCITTAGAAGAAAGATTGCCAGATGGTGCT----- 999
QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db :|||
999 ----- 999
QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
Db :|||
1000 -----ATCAGTGGAACTGGTTCTCAGAGTTTCTACAAATGCA 1035
QY 381 ArgIysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
Db :|||
1036 AAA----- 1038
QY 401 LysAspLeuProSerGluThrValIysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
Db :|||
1039 -----CCTAATCAAGTAGTGTCTAGCTAGGAGTCTTCAAGCAATCCTTCTTCT 1089
QY 421 ValSerHisThrLeuAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
Db :|||
1090 -----TTACGACAGTAAGAGCTCTCTTCAGCATCTGATGTTATATTTT 1137
QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db :|||
1138 AATCCAAAA---GATATCGTTGAAGAAACGGCTACAGCTTATATTCAGACATGGT--- 1191
QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
Db :|||
1192 ---GATCATTTTCATTACATTCAAAA-----TCAATCAAAATGGG 1230
QY 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db :|||
1231 CCAACGACTCTCCAAACAAATAGTCTAGCAACACCTTCTCCA---TCTCTTCCAATCAAT 1287
QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
Db :|||
1288 CCAGGAACCTTCATCAGAAACATGAA----- 1314
QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
Db :|||
1315 -----GAAGATGGATACGGATTGATGCTAATCGTATTATCGCTGAA 1356
QY 541 GlucylAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560
Db :|||
1357 GATCAATCAGGTTTGTGTCATGAGTCACGGAGACCAACATCATTATTTCTTCAAGAGGAC 1416
QY 561 LeuSerAspLysGlu 565
Db :|||
1417 TTGACAGAAGAGCAA 1431

```

90 7 0.9 104 17 Q8u029 pyrococcus

ALIGNMENTS

RESULT 1

Q9AHT9
ID Q9AHT9 PRELIMINARY; PRT; 816 AA.
AC Q9AHT9;
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE PHTA.
GN PHTA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
CX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N4;
RX MEDLINE=21116976; PubMed=11179332;
RA Wisemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,
RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tucmanet E.,
RA Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,
RA Langemann S., Johnson S., Koenig S.,
RT "Use of a Whole Genome Approach To Identify Vaccine Molecules
Affording Protection Against Streptococcus pneumoniae Infection.";
RL Infect. Immun. 69:1593-1598(2001).
DR EMBL: AF291695; AAK:9155.1; -
DR InterPro: IPR006270; Strep_his_triad.
DR Pfam: PF04270; strep_his_triad; 5.
DR TIGRFAMs: TIGR01363; strep_his_triad; 2.
SQ SEQUENCE 8-6 AA; 91519 MW; 5359126A611D27ED CRC64;

Query Match 99.9%; Score 995; DB 2; Length 816;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SYELGLYQARTVKNRRVSYIDGKQATOKTENLTPEVSKREGINAEQIVIKITDQGYVT 60
DB 21 SYELGLYQARTVKNRRVSYIDGKQATOKTENLTPEVSKREGINAEQIVIKITDQGYVT 80
QY 61 SHGDHYHYNGKVPYDAIISEELMKDPNYKLDKEDIIVNEVKGYYIKVQDKYVYVLKDA 120
DB 81 SHGDHYHYNGKVPYDAIISEELMKDPNYKLDKEDIIVNEVKGYYIKVQDKYVYVLKDA 140
QY 121 AHADNVRTKEEINRQKQEHSHQREGGTPRNDGVALARSQGYTTDGYIFNASDIIEDT 180
DB 141 AHADNVRTKEEINRQKQEHSHQREGGTPRNDGVALARSQGYTTDGYIFNASDIIEDT 200
QY 181 GDYIIVPHGDHYHYIPKNELASLAAAEAFLSGRGNLSNRYRQNSDNTSRTNWVPS 240
DB 201 GDYIIVPHGDHYHYIPKNELASLAAAEAFLSGRGNLSNRYRQNSDNTSRTNWVPS 260
QY 241 VSNPGTNTNTNSNTNSQASQNSNDISLLKOLYKPLSQRHVESDGLVDPQAITSRT 300
DB 261 VSNPGTNTNTNSNTNSQASQNSNDISLLKOLYKPLSQRHVESDGLVDPQAITSRT 320
QY 301 ARGVAVPHGDHYHYIPYQMSLELEERARIIPLYRSNHWVDSRPEQSPQPTPEPSPG 360
DB 321 ARGVAVPHGDHYHYIPYQMSLELEERARIIPLYRSNHWVDSRPEQSPQPTPEPSPG 380
QY 361 PQAPNLKIDSSLSVLVRKVGEGYFEKGISRYVFAKLPSETVKNLESKLSKQBS 420
DB 381 PQAPNLKIDSSLSVLVRKVGEGYFEKGISRYVFAKLPSETVKNLESKLSKQBS 440
QY 421 VSHLTATKENVAPRQEPFYDAKYNLLTEAHKALFNKGRNSDFCALDKLERLNDESTN 480
DB 441 VSHLTATKENVAPRQEPFYDAKYNLLTEAHKALFNKGRNSDFCALDKLERLNDESTN 500
QY 481 KEKLVDDLALFLAPITHPERLKGKPSQIEYTEDVRIQLADKYTSDGYIFDEHDIISD 540

DB 501 KEKLVDDLALFLAPITHPERLKGKPSQIEYTEDVRIQLADKYTSDGYIFDEHDIISD 560
QY 541 EGDAYVTPHGHSHWIGKQSLSKKEKVAAQAYTKGKILPSPDADVKANFTGSAAYI 600
DB 561 EGDAYVTPHGHSHWIGKQSLSKKEKVAAQAYTKGKILPSPDADVKANFTGSAAYI 620
QY 601 NRVKGEKRIPLVRLPYVMVHTVEVNGNLIIPKOHYHNIKFAWFDHDTYKAPNGYILED 660
DB 621 NRVKGEKRIPLVRLPYVMVHTVEVNGNLIIPKOHYHNIKFAWFDHDTYKAPNGYILED 680
QY 661 LFATIKYVVEHPDERPHSNDGWNASEHVLGKDHSDPNKNFKADBEPEETPAEPEVP 720
DB 681 LFATIKYVVEHPDERPHSNDGWNASEHVLGKDHSDPNKNFKADBEPEETPAEPEVP 740
QY 721 QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLNLTQIMDNNSIMAEAEKLLA 780
DB 741 QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLNLTQIMDNNSIMAEAEKLLA 800
QY 781 LLKGSNPSSVSKEKIN 796
DB 801 LLKGSNPSSVSKEKIN 816

RESULT 2

Q97QM8
ID Q97QM8 PRELIMINARY; PRT; 802 AA.
AC Q97QM8;
DT 01-OCT-2001 (TREMELrel. 18, Created)
DT 01-OCT-2001 (TREMELrel. 18, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Conserved domain protein.
GN SP1175.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
CX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson M.C., Peterson J.D.,
RA Mayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapfle E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson I., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL: AE007418; AAK75284.1; -
DR PIR: C95136; C95136.
DR TIGR: SP1175; -
DR InterPro: IPR006270; Strep_his_triad.
DR Pfam: PF04270; strep_his_triad; 5.
DR TIGRFAMs: TIGR01363; strep_his_triad; 2.
KW Complete proteome.
SQ SEQUENCE 802 AA; 90080 MW; 4E5CB8364EEA1933 CRC64;

Query Match 87.3%; Score 695; DB 16; Length 802;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKNRRVSYIDGKQATOKTENLTPEVSKREGINAEQIVIKITDQGYVT 60
DB 7 SYELGLYQARTVKNRRVSYIDGKQATOKTENLTPEVSKREGINAEQIVIKITDQGYVT 66
QY 61 SHGDHYHYNGKVPYDAIISEELMKDPNYKLDKEDIIVNEVKGYYIKVQDKYVYVLKDA 120
DB 67 SHGDHYHYNGKVPYDAIISEELMKDPNYKLDKEDIIVNEVKGYYIKVQDKYVYVLKDA 126

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 07:24:16 ; Search time 78 Seconds
(without alignments)
3219.904 Million cell updates/sec

Title: US-09-765-271-56
Perfect score: 796
Sequence: 1 SYELGYQARTVKNRVS.....KLLALLKGSNESSVSKKIN 796

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 10-7041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mic.*
- 8: sp_organella.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirs.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	795	99.9	816	2	Q9AHT9
2	695	87.3	802	16	Q97QM8
3	616	77.4	828	16	Q8DPQ2
4	107	13.4	844	2	Q9AG74
5	107	13.4	855	16	Q8CWR4
6	60	7.5	819	2	Q9ANY3
7	60	7.5	819	16	Q97QM9
8	60	7.5	819	16	Q9ANY2
9	60	7.5	853	16	Q8DQ08
10	32	4.0	1033	16	Q9ANY1
11	22	4.0	1039	16	Q8DQ07
12	27	3.4	481	16	Q8E338
13	17	2.1	289	2	Q9AE21
14	17	2.1	794	16	Q877Y2
15	17	2.1	822	2	Q9ZHG7
16	17	2.1	822	16	Q8E4U1

17	17	2.1	822	16	Q8D281	Q8d281 streptococc
18	17	2.1	823	16	Q8N282	Q8n282 streptococc
19	17	2.1	823	16	Q8K5Q1	Q8k5q1 streptococc
20	17	2.1	825	2	Q93GT5	Q93gt5 streptococc
21	17	2.1	825	16	Q99XV4	Q99xv4 streptococc
22	10	1.3	182	16	Q8DQ06	Q8dq06 streptococc
23	23	1.3	250	16	Q7WNG8	Q7wng8 bordetella
24	10	1.3	250	16	Q7WB03	Q7wb03 bordetella
25	10	1.3	250	16	Q7VY29	Q7vy29 bordetella
26	9	1.1	791	16	Q9CL01	Q9cl01 pasteurilla
27	8	1.0	71	12	Q8B9M0	Q8b9m0 rachiplusia
28	8	1.0	117	16	Q89NM4	Q89nm4 brachyrihob
29	8	1.0	147	16	Q8EHD8	Q8ehd8 shewanella
30	8	1.0	152	5	Q8IIB9	Q8iib9 plasmodium
31	8	1.0	154	5	Q91742	Q91742 hydra atten
32	8	1.0	181	5	Q44055	Q44055 drosophila
33	8	1.0	270	5	Q9NCF0	Q9ncf0 cryptocercu
34	8	1.0	284	5	Q22866	Q22866 caenorhabdi
35	8	1.0	284	5	Q22865	Q22865 caenorhabdi
36	8	1.0	284	5	Q27284	Q27284 caenorhabdi
37	8	1.0	294	5	Q46010	Q46010 caenorhabdi
38	8	1.0	301	5	Q95Q31	Q95q31 caenorhabdi
39	8	1.0	319	16	Q8FLF7	Q8flf7 corynebacte
40	8	1.0	330	13	Q7ZWH1	Q7zwh1 brachydancio
41	8	1.0	400	16	Q82AN2	Q82av2 streptomyc
42	8	1.0	436	3	Q9P8N1	Q9p8n1 coriolus ve
43	8	1.0	484	16	Q9KS51	Q9ks51 oryza sativ
44	8	1.0	485	10	Q80401	Q80401 oryza sativ
45	8	1.0	488	16	Q99TD8	Q99td8 staphylococ
46	8	1.0	488	16	Q8NM40	Q8nmw2 staphylococ
47	8	1.0	490	16	Q8CNR2	Q8cnc2 staphylococ
48	8	1.0	673	6	Q95LH9	Q95lh9 oryctolagus
49	8	1.0	675	16	Q89MC4	Q89mc4 bradyrhizob
50	8	1.0	840	16	Q9RSJ4	Q9rsj4 deinococcus
51	8	1.0	945	16	Q26077	Q26077 helicobacte
52	8	1.0	1169	2	Q84EVA	Q84ev4 methylococc
53	8	1.0	1205	16	Q7TXM2	Q7txm2 mycobacteri
54	8	1.0	1233	17	Q8TMX4	Q8tmx4 methanosaar
55	8	1.0	1646	2	Q93HW0	Q93hw0 streptococc
56	8	1.0	1659	16	Q97S90	Q97s90 streptococc
57	8	1.0	1659	16	Q8CZ52	Q8cz52 streptococc
58	8	1.0	2621	5	Q8SSW5	Q8ssw5 dictyosteli
59	7	0.9	42	15	Q87936	Q87936 chimpanzee
60	7	0.9	42	15	Q87948	Q87948 chimpanzee
61	7	0.9	42	15	Q87938	Q87938 chimpanzee
62	7	0.9	42	15	Q88084	Q88084 chimpanzee
63	7	0.9	42	15	Q87952	Q87952 chimpanzee
64	7	0.9	42	15	Q87946	Q87946 chimpanzee
65	7	0.9	42	15	Q88080	Q88080 chimpanzee
66	7	0.9	43	15	Q87944	Q87944 chimpanzee
67	7	0.9	44	15	Q87942	Q87942 chimpanzee
68	7	0.9	49	15	Q87940	Q87940 chimpanzee
69	7	0.9	49	16	Q45596	Q45596 bacillus su
70	7	0.9	50	2	Q9AGM9	Q9agm9 clostridium
71	7	0.9	51	16	Q8PHV0	Q8phv0 escherichia
72	7	0.9	53	10	Q7XKX3	Q7xkx3 oryza sativ
73	7	0.9	68	15	Q90VA0	Q90va0 human immun
74	7	0.9	68	16	Q9AJ30	Q9aj30 caulobacter
75	7	0.9	71	1	Q977L2	Q977l2 uncultured
76	7	0.9	71	17	Q96YD8	Q96yd8 sulfolobus
77	7	0.9	80	3	Q9HGR7	Q9hgr7 gilbertella
78	7	0.9	81	16	Q83L13	Q83l13 shigella fl
79	7	0.9	83	5	Q86ID1	Q86id1 dictyosteli
80	7	0.9	86	15	Q73850	Q73850 human immun
81	7	0.9	86	15	Q73851	Q73851 human immun
82	7	0.9	86	15	Q73852	Q73852 human immun
83	7	0.9	90	17	Q97W98	Q97w98 sulfolobus
84	7	0.9	92	13	Q83KC6	Q83kc6 shigella fl
85	7	0.9	93	16	Q9RUG5	Q9rug5 deinococcus
86	7	0.9	94	16	Q9RY27	Q9ry27 deinococcus
87	7	0.9	96	16	Q8FAP6	Q8fap6 escherichia
88	7	0.9	98	2	Q7WV63	Q7wv63 bacillus su
89	7	0.9	98	16	Q7V2U9	Q7v2u9 prochloroco

This Page Blank (uspto)

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21137209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwin M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radzue D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Anguoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter C.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.,
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001);
DR EMBL; AF318956; AAK06761.1;
DR EMBL; AE007403; AAK75121.1;
DR PIR; H95115; H95115;
DR TIGR; SP1004;
DR InterPro; IPR006270; Strep_his_triad;
DR Pfam; PF04270; strep_his_triad; 6;
DR TIGRFAMs; TIGR01363; strep_his_triad; 4;
KW Signal, Hypothetical protein; Complete proteome.
FT SIGNAL 1 29
SQ SEQUENCE 1039 AA; 11463 MW; 81A563FC806625C4 CRC64;

Query Match 4.0%; Score 32; DB 16; Length 1039;
Best Local Similarity 100.0%; Pred. No. 5.9e-22;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 AEQIVIKITDQGYTSHGHDHYHYNGKVPYDA 77
Db AEQIVIKITDQGYTSHGHDHYHYNGKVPYDA 98

RESULT 11
Q8DQ07 PRELIMINARY; PRT; 1039 AA.
AC Q8DQ07; 2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Pneumococcal histidine triad protein E.
GN PHE OR SP0908
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6)
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
CC Streptococcus.
CX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arrold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.F., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushina P.,
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Morris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6."
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008464; AAK99712.1;
DR PIR; D97985; D97985;
DR InterPro; IPR006270; Strep_his_triad;
DR Pfam; PF04270; strep_his_triad; 6;
DR TIGRFAMs; TIGR01363; strep_his_triad; 3;
KW Complete proteome.
SQ SEQUENCE 1039 AA; 114625 MW; 05CC226D2028F551 CRC64;

Query Match 4.0%; Score 32; DB 16; Length 1039;
Best Local Similarity 100.0%; Pred. No. 5.9e-22;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 46 AEQIVIKITDQGYTSHGHDHYHYNGKVPYDA 77
Db AEQIVIKITDQGYTSHGHDHYHYNGKVPYDA 98

RESULT 12
Q8E338 PRELIMINARY; PRT; 481 AA.
AC Q8E338;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN GHS1925.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
CC Streptococcus.
CX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22242508; PubMed=12354221;
RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Zouine M., Cauve E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease."
RL Mol. Microbiol. 45:1499-1513(2002).
DR EMBL; AL766854; CAD47584.1;
DR Sagaliet; G81925;
DR InterPro; IPR006270; Strep_his_triad;
DR Pfam; PF04270; strep_his_triad; 4;
DR TIGRFAMs; TIGR01363; strep_his_triad; 1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 481 AA; 53325 MW; DEF0453D8A929BE CRC64;

Query Match 3.4%; Score 27; DB 16; Length 481;
Best Local Similarity 100.0%; Pred. No. 2.5e-17;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 GYVTSHGHDHYHYNGKVPYDAIISEEL 83
Db GYVTSHGHDHYHYNGKVPYDAIISEEL 107

RESULT 13
Q9AE21 PRELIMINARY; PRT; 289 AA.
AC Q9AE21;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (fragment).
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
CC Streptococcus.
CX NCBI_TaxID=1311;
RN [-]
RP SEQUENCE FROM N.A.
RX STRAIN=5531;
RX MEDLINE=21172873; PubMed=11274116;
RA Granlund M., Mäkelä E., Norgren M.;
RT "Mutually exclusive distribution of IS1548 and CBS11, an active group
RT II intron identified in human isolates of group B streptococci."
RL J. Bacteriol. 183:2560-2569(2001).
DR EMBL; AJ290952; CAC35985.1;
DR InterPro; IPR006270; Strep_his_triad;
DR Pfam; PF04270; strep_his_triad; 2;
DR TIGRFAMs; TIGR01363; strep_his_triad; 1;
KW Hypothetical protein.
FT NON TER 289
SQ SEQUENCE 289 AA; 32043 MW; A15A8588EA8140E4 CRC64;

```

This Page Blank (uspto)


```

RT pneumoniae";
RL Science 293:498-506(2001);
DR EMBL; A5007418; AAK75283.1; -;
DR PIR; B95136; B95136;
DR TIGR; SP1174; -;
DR InterPro; IPR006270; Strep_his_triad;
DR Pfam; PF04270; strep_his_triad; 5;
DR TIGRFAMs; TIGR01363; strep_his_triad; 2;
KW Complete proteome.
SQ SEQUENCE 819 AA; 92228 MW; 43852B72E9163BDE CRC64;

Query Match
Best Local Similarity 7.5%; Score 60; DB 15; Length 819;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

31 ENLTPDEVSKREGINAEQIVIKITDQGYTSHGDHYYNGKVPYDAIISELLMKDPNY 90
52 ENLTPDEVSKREGINAEQIVIKITDQGYTSHGDHYYNGKVPYDAIISELLMKDPNY 111

RESULT 8
Q9ANY2 PRELIMINARY; PRT; 839 AA.
AC Q9ANY2
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 24, Last annotation update)
DE Pneumococcal histidine triad protein D precursor (Hypothetical protein
DE SPI003) (Fragment).
GN PHTD OR SPI003.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J.H., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angluoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae";
RL Science 293:498-506(2001);
DR EMBL; AF318955; AAK06760.1; -;
DR EMBL; AEC07403; AAK75120.1; -;
DR PIR; G95115; G95115;
DR TIGR; SPI003; -;
DR InterPro; IPR006270; Strep_his_triad;
DR Pfam; PF04270; strep_his_triad; 5;
DR TIGRFAMs; TIGR01363; strep_his_triad; 2;
KW SIGNAL, Hypothetical protein; Complete proteome.
FT SIGNAL 1 29
FT NON_TER 839 839
SQ SEQUENCE 839 AA; 93672 MW; 713B180D5E03BDCA CRC64;

Query Match
Best Local Similarity 7.5%; Score 60; DB 15; Length 839;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

31 ENLTPDEVSKREGINAEQIVIKITDQGYTSHGDHYYNGKVPYDAIISELLMKDPNY 90
52 ENLTPDEVSKREGINAEQIVIKITDQGYTSHGDHYYNGKVPYDAIISELLMKDPNY 111

RESULT 9
Q8DQ08 PRELIMINARY; PRT; 853 AA.
AC Q8DQ08
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Pneumococcal histidine triad protein D.
GN PHTD OR SPK0907.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.S., Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.V., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsuoka P.,
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Nicase T.L.,
RA Norris P.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R., Jr., Skatrud P.L.,
RA Glas J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008464; AAK99711; -;
DR PIR; C97985; C97985;
DR InterPro; IPR006270; Strep_his_triad;
DR Pfam; PF04270; strep_his_triad; 5;
DR TIGRFAMs; TIGR01363; strep_his_triad; 2;
KW Complete proteome.
SQ SEQUENCE 853 AA; 95225 MW; 9BF06A1EDE990CF5 CRC64;

Query Match
Best Local Similarity 7.5%; Score 60; DB 15; Length 853;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

31 ENLTPDEVSKREGINAEQIVIKITDQGYTSHGDHYYNGKVPYDAIISELLMKDPNY 90
52 ENLTPDEVSKREGINAEQIVIKITDQGYTSHGDHYYNGKVPYDAIISELLMKDPNY 111

RESULT 10
Q9ANY1 PRELIMINARY; PRT; 1039 AA.
AC Q9ANY1
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 24, Last annotation update)
DE Pneumococcal histidine triad protein E precursor (Hypothetical protein
DE SPI004).
GN PHTD OR SPI004.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21101045; PubMed=11159990;
RA Adamou J.B., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Dormitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
RA Langermann S., Koenig S., Johnson S.;
RT "Identification and characterization of a novel family of pneumococcal
RT proteins (the Pht family) that are protective against sepsis.";
RL Infect. Immun. 69:949-958(2001).

```

This Page Blank (uspto)

QY	121	AHADNVRTKEINRQKEHSQHREGGTPRNDCAVALARSQGRYTTDDGYIFNASDIISCT	180
Db	127	AHADNVRTKEINRQKEHSQHREGGTPRNDCAVALARSQRYTDDGYIFNASDIISBT	186
QY	181	GDAYIVPHGDHYHYPKNELASELAAAAFISGRGNLSNRTYRQNSDNDSRTNWVPS	240
Db	187	GDAYIVPHGDHYHYPKNELASELAAAAFISGRGNLSNRTYRQNSDNDSRTNWVPS	246
QY	241	VSNPGTNTNTNNSNTNSQAQSQNDISLLKQLYKLPLSQRHVESDGLVFPDPAITSR	300
Db	247	VSNPGTNTNTNNSNTNSQAQSQNDISLLKQLYKLPLSQRHVESDGLVFPDPAITSR	305
QY	301	ARGVAVEHCHYHFPYSQMSLEBERIARIIPRYRSNHWVPDSPEQSPQTPESPSPG	360
Db	307	ARGVAVEHCHYHFPYSQMSLEBERIARIIPRYRSNHWVPDSPEQSPQTPESPSPG	366
QY	361	POPAPNLKIDNSSVSQLVRKVGYYVFEKGISRYVFXADLPESETVKNLESKLSQES	420
Db	367	POPAPNLKIDNSSVSQLVRKVGYYVFEKGISRYVFXADLPESETVKNLESKLSQES	426
QY	421	VSHTLTAKENVAERDQFYDKAYNLLTEAHKALFXNKGNSDFQALDKLERLNDSETN	480
Db	427	VSHTLTAKENVAERDQFYDKAYNLLTEAHKALFXNKGNSDFQALDKLERLNDSETN	485
QY	481	KEKVDLLAFIAPITHPERLGKNSQIEYDEDEVRIAQLADKYYTSDGYIFUEHDIISD	540
Db	487	KEKVDLLAFIAPITHPERLGKNSQIEYDEDEVRIAQLADKYYTSDGYIFUEHDIISD	546
QY	541	EGDAYVTPHMGHSWIKGKSLDKKEVAAQAAYTKEKGIPLPSPDADVKANPTGDSAAIY	600
Db	547	EGDAYVTPHMGHSWIKGKSLDKKEVAAQAAYTKEKGIPLPSPDADVKANPTGDSAAIY	606
QY	601	NRVKGKRIPLVRLPYWVEHTVYKGNLIIIPKHQYHNTEKPAWFDCHTYKAPNGVTLED	660
Db	607	NRVKGKRIPLVRLPYWVEHTVYKGNLIIIPKHQYHNTEKPAWFDCHTYKAPNGVTLED	666
QY	661	LFATIKYYVEHPDERPHSNDGWNASHVLGKXHDSDPNKNTKADSEPVETPARPVP	720
Db	667	LFATIKYYVEHPDERPHSNDGWNASHVLGKXHDSDPNKNTKADSEPVETPARPVP	726
QY	721	QVETEKVEAQLKEAEVLLAKVTSSILKANATETLAGLRNLTLOIYDNNNSIMAEAPKLLA	780
Db	727	QVETEKVEAQLKEAEVLLAKVTSSILKANATETLAGLRNLTLOIYDNNNSIMAEAPKLLA	786
QY	781	LLKGSNPSSVSKIN	796
Db	787	LLKGSNPSSVSKIN	802

RESULT 3

Q8DPQ2	PRELIMINARY; PRT; 828 AA.
ID	
Q8DPQ2	
AC	
Q8DPQ2	
DT	01-VAR-2003 (TEMBLrel. 23, Created)
DT	01-VAR-2003 (TEMBLrel. 23, Last sequence update)
DT	01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE	Pneumococcal histidine triad protein A.
DE	PH1A OR SPRI361.
GN	Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OS	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC	Streptococcus.
OC	NCBI_TaxID=171101;
OX	[1]
OX	SEQUENCE FROM N.A.
RP	MEDLINE=21429245; PubMed=11544234;
RP	Hoskins J., Alborn W.E. Jr, Arnold J., Blaszcak L.C., BURGESS S.,
RA	DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA	Gilmour R., Glass C.S., Khoja H., Kraft A.R., Lagace R.E.,
RA	LeBlanc D.-J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA	McAnaran S.M., McHenry M., McLeester K., Mundy C.W., Nicas T.I.,
RA	Norris F.H., O'Garra M., Feery R.B., Robertson G.T., Rockey P.,
RA	Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA	Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr, Skatrud P.L.,

RESULT 4

Q9AG74	PRELIMINARY;	PRT;	844 AA.
ID	Q9AG74		
AC	Q9AG74		
DT	01-JUN-2001	(TREMUREL 17, Created)	
DT	01-JUN-2001	(TREMUREL 17, Last sequence update)	
DT	01-JUN-2003	(TREMUREL 24, Last annotation update)	
DE	PHPA.		
EN	PHPA.		

This Page Blank (uspto)

OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2124685; PubMed=11349048;
 RA Zhang Y., Masi A.W., Barniak V., Mountzouros K., Hostetter M.K.,
 RA Green B.A.;
 RT "Recombinant PnpA Protein, a Unique Histidine Motif-Containing Protein
 RT from Streptococcus pneumoniae, Protects Mice against Intranasal
 RL Pneumococcal Challenge";
 RL Infect. Immun. 69:13827-13836(2001).
 DR EMBL; AF340221; AAK26629.1; .
 DR InterPro; IPR006270; Strep_his_triad.
 DR Pfam; PF04270; strep_his_triad; 5
 DR TIGRFAMs; TIGR01363; strep_his_triad; 2.
 SQ SEQUENCE 844 AA; 94769 MW; 5738A52290FF8902 CRC64;
 Query Match 13.4%; Score 107; DB 2; Length 844;
 Best Local Similarity 130.0%; Pred. No. 7.1e-96;
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 102 KGGVIVKDGKYYVYLKDAHADNVRTKEEINRQKQHSQHREGTPRNDGAVALARSG 161
 Db 122 KGGVIVKDGKYYVYLKDAHADNVRTKEEINRQKQHSQHREGTPRNDGAVALARSG 181
 QY 162 RYTTDGGYIFNASDIETGDAYIVPHGDHVIYIPKNLSASELAAA 208
 Db 182 RYTTDGGYIFNASDIETGDAYIVPHGDHVIYIPKNLSASELAAA 228
 RESULT 5
 Q8CWR4
 ID Q8CWR4 PRELIMINARY; PRT; 855 AA.
 AC Q8CWR4;
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 DE Histidine motif-containing protein.
 GN PHPA OR SP1060.
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OX NCBI_TaxID=171101;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21429245; PubMed=11544234;
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
 RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
 RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
 RA LeBlanc D.J., Lee J.N., Lefkowitz E.J., Lu J., Matsushima P.,
 RA Mahren S.M., McMenney M., McLeaster K., Mundy C.W., Niclas T.I.,
 RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rocky P.,
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,
 RA Glass J.I.;
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6";
 RL J. Bacteriol. 183:5709-5717(2001).
 DR EMBL; AB008479; AAK99864.1; .
 DR PIR; D98004; D98004.
 DR InterPro; IPR006270; Strep_his_triad.
 DR Pfam; PF04270; strep_his_triad; 5.
 DR TIGRFAMs; TIGR01363; strep_his_triad; 2.
 SQ Complete proteome.
 QY SEQUENCE 855 AA; 96177 MW; 4350E82A3F97089A CRC64;
 Query Match 13.4%; Score 107; DB 16; Length 855;
 Best Local Similarity 100.0%; Pred. No. 7.2e-96;
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 102 KGGVIVKDGKYYVYLKDAHADNVRTKEEINRQKQHSQHREGTPRNDGAVALARSG 161
 Db 122 KGGVIVKDGKYYVYLKDAHADNVRTKEEINRQKQHSQHREGTPRNDGAVALARSG 181
 QY 162 RYTTDGGYIFNASDIETGDAYIVPHGDHVIYIPKNLSASELAAA 208
 Db 182 RYTTDGGYIFNASDIETGDAYIVPHGDHVIYIPKNLSASELAAA 228

Db 133 KGGVIVKDGKYYVYLKDAHADNVRTKEEINRQKQHSQHREGTPRNDGAVALARSG 192
 QY 162 RYTTDGGYIFNASDIETGDAYIVPHGDHVIYIPKNLSASELAAA 208
 Db 193 RYTTDGGYIFNASDIETGDAYIVPHGDHVIYIPKNLSASELAAA 239
 RESULT 6
 Q9ANY3
 ID Q9ANY3 PRELIMINARY; PRT; 819 AA.
 AC Q9ANY3;
 DT 01-JUN-2001 (T-EMBLrel. 17, Created)
 DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
 DE Pneumococcal histidine triad protein B precursor (fragment).
 GN PHTB.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21101045; PubMed=11159990;
 RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
 RA Dornitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
 RA Langermann S., Koenig S., Johnson S.;
 RT "Identification and Characterization of a novel family of pneumococcal
 RT proteins (the Pht family) that are protective against sepsis";
 RL Infect. Immun. 69:949-958(2001).
 DR EMBL; AF318954; AAK06759.1; .
 DR InterPro; IPR006270; Strep_his_triad.
 DR Pfam; PF04270; strep_his_triad; 5.
 DR TIGRFAMs; TIGR01363; strep_his_triad; 2
 KW SIGNAL
 FT NON_TER 819 819 POTENTIAL.
 SQ SEQUENCE 819 AA; 92109 MW; E602CFCL6CC28A5F CRC64;
 Query Match 7.5%; Score 60; DB 2; Length 819;
 Best Local Similarity 100.0%; Pred. No. 1.3e-49;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 31 ENLTDEVSKEGNAQIVIKITDQGYVTSAGDHVYGVYDAIISLLMKDPNY 90
 Db 52 ENLTDEVSKEGNAQIVIKITDQGYVTSAGDHVYGVYDAIISLLMKDPNY 111
 RESULT 7
 Q97QM9
 ID Q97QM9 PRELIMINARY; PRT; 819 AA.
 AC Q97QM9;
 DT 01-OCT-2001 (T-EMBLrel. 18, Created)
 DT 01-OCT-2001 (T-EMBLrel. 18, Last sequence update)
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
 DE Conserved domain protein.
 GN SP1174.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=ATCC BAA-334 / TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.S.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus

This Page Blank (uspto)

Query Match 2.1%; Score 17; DB 2; Length 289;
 Best Local Similarity 100.0%; Pred. No. 1.1e-07;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YNGKVPYDAIISELLM 85
 |||||
 Db 92 YNGKVPYDAIISELLM 108

RESULT 14

Q877Y2 PRELIMINARY; PRT; 794 AA.
 AC Q877Y2;
 DT 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN SP51722.
 OS Streptococcus pyogenes (serotype M3).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=198466;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SSI-1 / Serotype M3;
 RA Nakagawa I., Kurokawa K., Nakata M., Tomiyasu Y., Yamashita A.,
 RA Yamazaki K., Okahashi N., Kawabata S., Yasunaga T., Hattori M.,
 RA Hayashi H., Hamada S.;
 RT "The genome of invasive Streptococcus pyogenes; a comparative analysis
 of S. pyogenes SSI-1, SF370 and MGAS8232."
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP005146; BAC64817.1;
 DR InterPro: IPR006270; Strep_his_triad.
 DR Pfam: PF04270; strep_his_triad; 6.
 DR Hypothetical protein.
 KW Hypothetical protein.
 SQ SEQUENCE 794 AA; 9B07269C8F5CB875 CRC64;

Query Match 2.1%; Score 17; DB 16; Length 794;
 Best Local Similarity 100.0%; Pred. No. 2.7e-07;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YNGKVPYDAIISELLM 85
 |||||
 Db 63 YNGKVPYDAIISELLM 79

RESULT 15

Q9ZHG7 PRELIMINARY; PRT; 822 AA.
 AC Q9ZHG7;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Streptococcus agalactiae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1311;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R268;
 RX MEDLINE=99115568; PubMed=9916102;
 RA Spellerberg B., Rodzinski E., Martin S., Weber-Heynemann J.,
 RA Schnitzler N., Luetticken R., Podbielski A.;
 RT "Imb, a protein with similarities to the Irai adhesin family, mediates
 attachment of Streptococcus agalactiae to human laminin."
 RL Infect. Immun. 67:811-818(1999).
 DR EMBL: AF062533; AAD13797.1;
 DR FTR: T46758; T46758.
 DR InterPro: IPR006270; Strep_his_triad.
 DR Pfam: PF04270; strep_his_triad; 6.
 DR TIGRFAMS: TIGR01363; strep_his_triad; 4.

KW Hypothetical protein.
 SQ SEQUENCE 822 AA; 92386 MW; 80E4EDF313491F98 CRC64;

Query Match 2.1%; Score 17; DB 2; Length 822;
 Best Local Similarity 100.0%; Pred. No. 2.8e-07;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YNGKVPYDAIISELLM 85
 |||||
 Db 92 YNGKVPYDAIISELLM 108

RESULT 16

Q8E4U1 PRELIMINARY; PRT; 822 AA.
 AC Q8E4U1;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN GBS1306.
 OS Streptococcus agalactiae (serotype III).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=216495;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NEM316 / Serotype III;
 RX MEDLINE=22242508; PubMed=12354221;
 RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Prangeul L.,
 RA Maadek T., Zouine M., Couve E., Lailoui L., Foyat C., Trieu-Cuot P.,
 RA Kunst F.;
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
 RT invasive neonatal disease."
 RL Mol. Microbiol. 45:1499-1513(2002).
 DR EMBL: AL766850; CAD46965.1;
 DR Sagalid; gbs1306;
 DR InterPro: IPR006270; Strep_his_triad.
 DR Pfam: PF04270; strep_his_triad; 6.
 DR TIGRFAMS: TIGR01363; strep_his_triad; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 822 AA; 92393 MW; 2929A97C8AFCD78F CRC64;

Query Match 2.1%; Score 17; DB 16; Length 822;
 Best Local Similarity 100.0%; Pred. No. 2.8e-07;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YNGKVPYDAIISELLM 85
 |||||
 Db 92 YNGKVPYDAIISELLM 108

RESULT 17

Q8DZ81 PRELIMINARY; PRT; 822 AA.
 AC Q8DZ81;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Streptococcal histidine triad family protein.
 GN SAG1233.
 OS Streptococcus agalactiae (serotype V).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=216466;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2603 V/R / Serotype V;
 RX MEDLINE=22222988; PubMed=12200547;
 RA Tettelin H., Massignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
 RA Wessels M.R., Paulsen I.T., Nelson K.E., Margalit I., Read T.D.,
 RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
 RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,

RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
RA Carty H.A., Cline K.T., Van Aken S.E., Gll J., Scarcelli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
RL EMBL; AE014248; AAN00111.1; -;
DR TIGR; SAG1233; -;
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 6.
DR TIGRFAMs; TIGR01363; strep_his_triad; 1.
KW Complete proteome.
SQ SEQUENCE 822 AA; 92400 MW; 8CC8DF316727F98 CRC64;
Query Match 2.1%; Score 17; DB 16; Length 822;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 69 YNGKVPYDAIISEELM 85
|||||
Db 92 YNGKVPYDAIISEELM 108
RESULT 18
Q8N282 PRELIMINARY; PRT; 823 AA.
AC Q8N282;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein spym18_2072.
GN SPYM18_2072.
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=186103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
DR EMBL; AE010110; AAL98543.1; -;
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 6.
DR TIGRFAMs; TIGR01363; strep_his_triad; 4.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 823 AA; 92585 MW; C7951EB30CEEDF0C CRC64;
Query Match 2.1%; Score 17; DB 16; Length 823;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 69 YNGKVPYDAIISEELM 85
|||||
Db 92 YNGKVPYDAIISEELM 108
RESULT 19
Q8K5Q1 PRELIMINARY; PRT; 823 AA.
AC Q8K5Q1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Histidine triad protein.
GN SPYM3_1724.
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=198466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=2213808; PubMed=1212206;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
RL EMBL; AE014169; AAM80331.1; -;
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 6.
DR TIGRFAMs; TIGR01363; strep_his_triad; 1.
KW Complete proteome.
SQ SEQUENCE 823 AA; 92431 MW; 8C6CBD517A2DD616 CRC64;
Query Match 2.1%; Score 17; DB 16; Length 823;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 69 YNGKVPYDAIISEELM 85
|||||
Db 92 YNGKVPYDAIISEELM 108
RESULT 20
Q93GT5 PRELIMINARY; PRT; 825 AA.
AC Q93GT5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Histidine triad protein of group A streptococci.
GN HTPA.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SSI-9;
RA Terao Y., Kawabata S., Hamada S.;
RT "Characterization of a novel histidine triad protein of group A
RT streptococci.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBSJ databases.
RL EMBL; AB073859; BAB71774.1; -;
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 6.
DR TIGRFAMs; TIGR01363; strep_his_triad; 4.
DR TIGR01363; strep_his_triad; 4.
SQ SEQUENCE 825 AA; 92623 MW; DE4ECC199181DFFB CRC64;
Query Match 2.1%; Score 17; DB 2; Length 825;
Best Local Similarity 100.0%; Pred. No. 2.8e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 69 YNGKVPYDAIISEELM 85
|||||
Db 92 YNGKVPYDAIISEELM 108
RESULT 21
Q99XV4 PRELIMINARY; PRT; 825 AA.
ID Q99XV4
AC Q99XV4;

RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.",
RL Nat. Genet. 35:32-40(2003).
DR EMBL: BX640426; CAE36508.1; --
KW ATP-binding; Complete proteome.
SQ SEQUENCE 250 AA; 27826 MW; 4D6D99723660A7F5 CRC64;

Query Match 1.3%; Score 10; DB 16; Length 250;
Best Local Similarity 100.0%; Pred.No. 0.76;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 515 VRIQLADKY 524
| | | | | | | | | |
Db 133 VRIQLADKY 142

RESULT 25
Q7VY29 PRELIMINARY; PRT; 250 AA.
AC Q7VY29;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative ATP-binding component of ABC transporter.
GN BP1534.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebatia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Fellwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.",
RL Nat. Genet. 35:32-40(2003).
DR EMBL: BX640415; CAE41823.1; --
KW ATP-binding; Complete proteome.
SQ SEQUENCE 250 AA; 27928 MW; 3EDAFD80C5BECCCA CRC64;

Query Match 1.3%; Score 10; DB 16; Length 250;
Best Local Similarity 100.0%; Pred.No. 0.76;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 515 VRIQLADKY 524
| | | | | | | | | |
Db 133 VRIQLADKY 142